QC 1
.U58

Volume 64B

Number 2

Apr.–June 1960

U. OF MICHIGAN
JUL 12 1960
PHYSICS
LIBRARY

Journal of Research

of the

NATIONAL BUREAU OF STANDARDS

B. MATHEMATICS AND MATHEMATICAL PHYSICS

Journal of Research
of the
National Bureau of Standards

B. Mathematics and Mathematical Physics

APR.–JUNE • 1960

VOLUME • 64B

NUMBER • 2

Editor: Chester H. Page

Associate Editors: F. L. Alt, E. L. Crow

Publication dates: Mar. 10, June 10, Sept. 10, Dec. 10, 1960.

JOURNAL OF RESEARCH

The National Bureau of Standards Journal of Research reports research and development in the fields of activity shown at right. Also included from time to time are survey articles on topics closely related to the Bureau's scientific and technical program.

The Journal is published in four separate sections as follows:

A. Physics and Chemistry

Contains papers of interest primarily to scientists working in these fields. Issued six times a year. Annual subscription: domestic, $4.00; foreign, $4.75.

B. Mathematics and Mathematical Physics

Presents studies and compilations designed mainly for the mathematician and the theoretical physicist. Issued quarterly. Annual subscription: domestic, $2.25; foreign, $2.75.

C. Engineering and Instrumentation

Reports research and development results of interest chiefly to the engineer and the applied scientist. Issued quarterly. Annual subscription: domestic, $2.25; foreign, $2.75.

D. Radio Propagation

Reports research in radio propagation, upper atmospheric physics, and communications. Issued six times a year. Annual subscription: domestic, $4.00; foreign, $4.75.

U.S. DEPARTMENT OF COMMERCE
Frederick H. Mueller, Secretary

NATIONAL BUREAU OF STANDARDS
A. V. Astin, Director

Functions and Activities

The functions of the National Bureau of Standards are set forth in the Act of Congress, March 3, 1901, as amended. These include the development and maintenance of the national standards of measurement and the provision of means and methods for making measurements consistent with these standards; the determination of physical constants and properties of materials; the development of methods and instruments for testing materials, devices, and structures; advisory services to government agencies on scientific and technical problems; invention and development of devices to serve special needs of the Government; and the development of standard practices, codes, and specifications. The work includes basic and applied research, development, engineering, instrumentation, testing, evaluation, calibration services, and various consultation and information services. The Bureau also serves as the Federal technical research center in a number of specialized fields. The scope of activities of the National Bureau of Standards is suggested in the following listing of the divisions and sections engaged in technical work.

Washington, D.C.

Electricity and Electronics. Resistance and Reactance. Electron Devices. Electrical Instruments. Magnetic Measurements. Dielectrics. Engineering Electronics. Electronic Instrumentation. Electrochemistry.

Optics and Metrology. Photometry and Colorimetry. Optical Instruments. Photographic Technology. Length. Engineering Metrology.

Heat. Temperature Physics. Thermodynamics. Cryogenic Physics. Rheology. Molecular Kinetics. Free Radicals Research.

Atomic and Radiation Physics. Spectroscopy. Radiometry. Mass Spectrometry. Solid State Physics. Electron Physics. Atomic Physics. Neutron Physics. Radiation Theory. Radioactivity. X-rays. High Energy Radiation. Nucleonic Instrumentation. Radiological Equipment.

Chemistry. Organic Coatings. Surface Chemistry. Organic Chemistry. Analytical Chemistry. Inorganic Chemistry. Electrodeposition. Molecular Structure and Properties of Gases. Physical Chemistry. Thermochemistry. Spectrochemistry. Pure Substances.

Mechanics. Sound. Mechanical Instruments. Fluid Mechanics. Engineering Mechanics. Mass and Scale. Capacity, Density, and Fluid Meters. Combustion Controls.

Organic and Fibrous Materials. Rubber. Textiles. Paper. Leather. Testing and Specifications. Polymer Structure. Plastics. Dental Research.

Metallurgy. Thermal Metallurgy. Chemical Metallurgy. Mechanical Metallurgy. Corrosion. Metal Physics.

Mineral Products. Engineering Ceramics. Glass. Refractories. Enameled Metals. Constitution and Microstructure.

Building Technology. Structural Engineering. Fire Protection. Air Conditioning, Heating, and Refrigeration. Floor, Roof, and Wall Coverings. Codes and Safety Standards. Heat Transfer. Concreting Materials.

Applied Mathematics. Numerical Analysis. Computation. Statistical Engineering. Mathematical Physics.

Data Processing Systems. SEAC Engineering Group. Components and Techniques. Digital Circuitry. Digital Systems. Analog Systems. Applications Engineering.

• Office of Basic Instrumentation.

• Office of Weights and Measures.

Boulder, Colorado

Cryogenic Engineering. Cryogenic Equipment. Cryogenic Processes. Properties of Materials. Gas Liquefaction.

Radio Propagation Physics. Upper Atmosphere Research. Ionosphere Research. Regular Prediction Services. Sun-Earth Relationships. VHF Research. Radio Warning Services. Airglow and Aurora. Radio Astronomy and Arctic Propagation.

Radio Propagation Engineering. Data Reduction Instrumentation. Radio Noise. Tropospheric Measurements. Tropospheric Analysis. Propagation-Terrain Effects. Radio-Meteorology. Lower Atmosphere Physics.

Radio Standards. High-Frequency Electrical Standards. Radio Broadcast Service. Radio and Microwave Materials. Electronic Calibration Center. Microwave Circuit Standards.

Radio Communication and Systems. Low Frequency and Very Low Frequency Research. High Frequency and Very High Frequency Research. Modulation Systems. Antenna Research. Navigation Systems. Systems Analysis. Field Operations.

Order all publications from the Superintendent of Documents
U.S. Government Printing Office, Washington 25, D.C.

Use of funds for printing this publication approved by the director of the Bureau of the Budget (June 24, 1958).

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 64B, No. 2, April–June 1960

Non-Self-Adjoint Boundary Value Problems in Ordinary Differential Equations

Werner Greub [1] and Werner C. Rheinboldt

(October 15, 1959)

It is shown that the theory of non-self-adjoint linear ordinary differential equations can be simplified and unified if, instead of specifying linear boundary conditions in the conventional way, one merely specifies the linear subspace determined by the boundary conditions. If this is done, the corresponding linear space of the adjoint problem is the orthogonal complement of the original space with respect to the scalar product defined by the right-hand side of Green's boundary formula.

1. Introduction

The theory of boundary-value problems for ordinary differential equations of the form

$$\sum_{\nu=0}^{n} a_{n-\nu}(t)u^{(\nu)}(t)=f(t) \tag{1}$$

is now regarded as more or less complete. The self-adjoint case is presented in most textbooks owing to its importance in applications. A complete treatment of the non-self-adjoint problem is less common but may be found in Coddington, Levinson,[2] and Ince.[3] However, in these as well as in the original publications, e.g., Bôcher,[4] the boundary conditions are stated in the form

$$\omega_\mu(u(a), \ldots, u^{(n-1)}(a), u(b), \ldots, u^{(n-1)}(b)) \equiv \sum_{\nu=0}^{n-1} (A_{\mu\nu}u^{(\nu)}(a)+B_{\mu\nu}u^{(\nu)}(b))=0 \qquad (\mu=1, \ldots, m) \tag{2}$$

where the ω_μ are m linearly independent linear functions in the $2n$ dimensional number space R^{2n} $(0 \leqq m \leqq 2n)$.

The m linear equations $\omega_\mu=0$ $(\mu=1, \ldots, m)$ define a $(2n-m)$ dimensional linear subspace $\Omega \in R^{2n}$. The set of solutions $u(t)$ of the problem (1)/(2) remains unchanged if the functions ω_μ are replaced by another system of equations $\overline{\omega}_\mu=0$ $(\mu=1, \ldots, m)$ which define the same subspace $\Omega \in R^{2n}$.

This fact causes an unnecessary complication for the theoretical investigation of such boundary value problems. It results especially in a somewhat artificial definition of the adjoint problem. These difficulties do not occur if one prescribes instead of the boundary conditions (2) only the subspace Ω of R^{2n}, disregarding the manner in which Ω is represented. If this is done, the corresponding linear space of the adjoint problem is the orthogonal complement of the original space Ω with respect to a scalar product defined by the right-hand side of Green's boundary value formula.

It is shown in this article that the theory of the non-self-adjoint problems can be simplified and unified by a consequent use of these ideas and the methods of linear algebra.

2. Green's Formula

As usual we denote by C^k $(k=0,1, \ldots)$ the space of all real-valued[5] functions $u(t)$ which possess k continuous derivatives on a given closed interval $[a,b]$. Let $L[u]$ be the linear differential operator of order $n(\geqq 1)$ defined by

$$L[u] \equiv \sum_{\nu=1}^{n} a_{n-\nu}(t)u^{(\nu)}(t) \tag{3}$$

where $a_{n-\nu}(t) \in C^\nu$, $u(t) \in C^n$ and $a_0(t) \neq 0$ for $[a,b]$. To every function $u(t) \in C^{n-1}$ one can compute the $2n$ real numbers

$$u(a), u'(a), \ldots, u^{(n-1)}(a), u(b), u'(b), \ldots, u^{(n-1)}(b). \tag{4}$$

We regard these $2n$ numbers as the components of a vector $\mathbf{u}$ of the $2n$-dimensional linear number space R^{2n} and call it the boundary value vector of $u(t)$. The correspondence between the function $u(t)$ and its boundary value vector $\mathbf{u}$ is a linear mapping

[1] The work was carried out while both authors were employed at the University of Maryland, the first in the Department of Mathematics, the second in the Institute for Fluid Dynamics and Applied Mathematics. The paper was prepared while both authors were affiliated with the National Bureau of Standards, the first working under a National Bureau of Standards contract with The American University. The authors' present addresses are, respectively: Mathematics Institute, University of Zurich, Zurich, Switzerland; and Computing Center, Syracuse University, Syracuse, N.Y.

[2] Earl A. Coddington, Norman Levinson, Theory of ordinary differential equations (McGraw-Hill Book Co., Inc., New York, N.Y., 1955).

[3] E. L. Ince, Ordinary differential equations, London (1927).

[4] M. Bôcher, Applications and generalizations of the conception of adjoint systems, Trans. Am. Math. Soc. **14**, 403 to 420 (1913).

[5] The restriction to real-valued functions is not essential, as can be readily ascertained.

$$\mathbf{u}=\rho(u) \qquad (u(t)\in C^{n-1}, \quad \mathbf{u}\in R^{2n}) \tag{5}$$

from C^{n-1} onto R^{2n}.

Let Ω be a given subspace of the R^{2n}. Then the boundary value problem in consideration assumes the form

$$L[u]=f(t), \qquad \rho(u)\in\Omega, \tag{π_N}$$

with given $f(t)\in C^0$. The corresponding homogeneous problem is

$$L[u]=0, \qquad \rho(u)\in\Omega. \tag{π_H}$$

For any two functions $u(t)$, $v(t)\in C^n$, the integral

$$(L[u], v)=\int_a^b L[u]\,v\,dt \tag{6}$$

is defined and by repeated partial integrations can be written as

$$(L[u], v)=(u, L^*[v])+[u,v]_a^b, \tag{7}$$

where

$$L^*[v]=\sum_{\nu=0}^{n}(-1)^\nu(a_{n-\nu})^{(\nu)}, \tag{8}$$

$$[u,v]_a^b=\sum_{\nu=0}^{n-1}\sum_{\mu=\nu+1}^{n}(-1)^{\mu-\nu-1}u^{(\nu)}(a_{n-\nu}v)^{\mu-\nu-1}\Big|_a^b. \tag{9}$$

Equation (7) is the well-known Green's formula.

The linear differential operator L^* is called the adjoint operator of L. The expression $[u,v]_a^b$ is a bilinear form in the two boundary value vectors $\rho(u)$, $\rho(v)$. For arbitrary vectors

$$\mathbf{x}=\{..,\xi_\nu^{(a)},..;..,\xi_\nu^{(b)},..\}, \qquad \mathbf{y}=\{..,\eta_\nu^{(a)},..;..,\eta_\nu^{(b)},..\}$$

of the R^{2n}, this bilinear form can be written in the form

$$\varphi(\mathbf{x},\mathbf{y})=\sum_{\nu=0}^{n-1}\sum_{\mu=0}^{n-\nu-1}\sum_{\lambda=\mu}^{n-\nu-1}(-1)^\lambda\binom{\lambda}{\mu}$$
$$\cdot[a_{n-\nu-\lambda-1}^{(\lambda-\mu)}(b)\xi_\nu^{(b)}\eta_\mu^{(b)}-a_{n-\nu-\lambda-1}^{(\lambda-\mu)}(a)\xi_\nu^{(a)}\eta_\mu^{(a)}]. \tag{10}$$

With this notation we get Green's formula,

$$(L[u], v)=(u, L^*[v])+\varphi(\rho(u), \rho(v)). \tag{11}$$

It is obvious that the correspondence $L\to L^*$ is linear. Furthermore, the adjoint operator to L^* is again L. To show this, apply Green's formula to any $u,v\in C^n$ with $\rho(u)=\rho(v)=0$:

$$(L[u], v)-(u, L^*[v])=0$$

$$(L^*[v], u)-(v, L^{**}[u])=0,$$

which results in

$$(L[u]-L^{**}[u], v)=0$$

(for all $u,v\in C^n$ with $\rho(u)=\rho(v)=0$) and thus leads to

$$L[u]=L^{**}[u]$$

for all functions $u(t)\in C^n$. In the last step one has to use the same technique used in calculus of variations for the proof of Euler's condition.

The differential operator L is said to be self-adjoint if $L^*\equiv L$. On comparing the forms (3) and (8) of L and L^* one sees immediately that n must be even if the operator is self-adjoint.

In the case of a self-adjoint operator L the left-hand side of (11) is skew-symmetric with respect to u and v, and this implies that the bilinear form $\varphi(\mathbf{u},\mathbf{v})$ is also skew-symmetric.

We will now show that $\varphi(\mathbf{x},\mathbf{y})$ is not degenerate. To prove this it is sufficient to verify the following statement: If for a fixed vector $\mathbf{y}\in R^{2n}$ we have $\varphi(\mathbf{x},\mathbf{y})=0$ for every $\mathbf{x}\in R^{2n}$, then it follows that $\mathbf{y}=0$.

We prove this first for a fixed vector $\mathbf{y}$ of the form

$$\mathbf{y}_a=\{\eta_0^{(a)}, \ldots, \eta_{n-1}^{(a)}, 0, \ldots, 0\} \tag{12a}$$

or

$$\mathbf{y}_b=\{0, \ldots, 0, \eta_0^{(b)}, \ldots, \eta_{n-1}^{(b)}\} \tag{12b}$$

taking x correspondingly as $\mathbf{x}_a$ or $\mathbf{x}_b$, analogously defined. In this case the expression $\varphi(\mathbf{x}_a, \mathbf{y}_a)$ for $\alpha=a$ or $\alpha=b$ is a bilinear form in the two n-dimensional vectors $\{\xi_0^{(\alpha)}, \ldots, \xi_{n-1}^{(\alpha)}\}$ and $\{\eta_0^{(\alpha)}, \ldots, \eta_{n-1}^{(\alpha)}\}$. Therefore, if $\varphi(\mathbf{x}_a, \mathbf{y}_a)=0$ holds for every $\mathbf{x}_a$ and a fixed $\mathbf{y}_a$, then in expression (10) the coefficient of every $\xi_\nu^{(\alpha)}$ must vanish separately. Thus

$$\sum_{\mu=0}^{n-\nu-1}\sum_{\lambda=\mu}^{n-\nu-1}(-1)^\lambda\binom{\lambda}{\mu}a_{n-\nu-\lambda-1}^{(\lambda-\mu)}(\alpha)\,\eta_\mu^{(\alpha)}=0$$
$$(\nu=0, \ldots, n-1).$$

For $\nu=n-1$ this has the form $a_0(\alpha)\,\eta_0^{(\alpha)}=0$, and since $a_0(\alpha)\neq 0$ it follows that

$$\eta_0^{(\alpha)}=0.$$

For $\eta=n-2$ we get now

$$a_1(\alpha)\,\eta_0^{(\alpha)}-a_0'(\alpha)\,\eta_0^{(\alpha)}-a_0(\alpha)\,\eta_1^{(\alpha)}=-a_0(\alpha)\,\eta_1^{(\alpha)}=0$$

and

$$\eta_1^{(\alpha)}=0.$$

Continuing in this way, we ultimately show that

$$\eta_\mu^{(\alpha)}=0 \qquad (\mu=0, \ldots, n-1)$$

and thus $\mathbf{y}_a=0$.

Now the proof is readily completed. From definition (10) it is seen that for arbitrary $\mathbf{x}=\mathbf{x}_a+\mathbf{x}_b$ and $\mathbf{y}=\mathbf{y}_a+\mathbf{y}_b$,

$$\varphi(\mathbf{x}, \mathbf{y})=\varphi(\mathbf{x}_b, \mathbf{y}_b)-\varphi(\mathbf{x}_a, \mathbf{y}_a).$$

For a fixed $\mathbf{y}$ this expression shall be zero for every $\mathbf{x} \in R^{2n}$ and in particular for arbitrary vectors $\mathbf{x}=\mathbf{x}_a$ or $\mathbf{x}=\mathbf{x}_b$. Hence the two terms $\varphi(\mathbf{x}_a, \mathbf{y}_a)$ and $\varphi(\mathbf{x}_b, \mathbf{y}_b)$ have to vanish separately for every $\mathbf{x}_a$ and $\mathbf{x}_b$, respectively. In accordance with the first part of the proof, this is only possible for $\mathbf{y}_a=0$ and $\mathbf{y}_b=0$, i.e., $\mathbf{y}=0$.

3. The Adjoint Problem

The bilinear form φ was shown to be nondegenerate. Accordingly, it can be employed to define "orthogonality" of vectors $\mathbf{x}$, $\mathbf{y} \in R^{2n}$. We will say $\mathbf{x}$ and $\mathbf{y}$ are φ-orthogonal if $\varphi(\mathbf{x}, \mathbf{y})=0$. It should be observed that this φ-orthogonality is not necessarily symmetrical.

One can speak, in particular, of the φ-orthogonal space to Ω that is the space $\Omega^{\perp}$ of all vectors $\mathbf{y} \in R^{2n}$ which are φ-orthogonal to all vectors $\mathbf{x} \in \Omega$.

For the dimension of $\Omega^{\perp}$ evidently holds

$$\dim \Omega^{\perp}=2n-\dim \Omega=m. \tag{13}$$

Equation (13) implies

$$\dim (\Omega^{\perp})^{\perp}=\dim \Omega,$$

and hence, together with $(\Omega^{\perp})^{\perp} \in \Omega$, that

$$(\Omega^{\perp})^{\perp}=\Omega. \tag{14}$$

It is worth while pointing out that the relation $\Omega \cap \Omega^{\perp}=0$ is not, in general, valid; there may exist a nonzero vector $\mathbf{x} \in \Omega$ which is φ-orthogonal to all vectors of Ω.

With this terminology we define the adjoint boundary value problem of (π_N):

$$(\pi_N^*) \qquad L^*[u]=g(t), \qquad \rho(u) \in \Omega$$

with given $g(t) \in C^0$.

From $L^{**}=L$ and (14), it follows that the adjoint problem of (π_N^*) is again (π_N).

In particular, (π_N) will be termed self-adjoint [6] if

$$L^* \equiv L, \qquad \Omega=\Omega^{\perp}. \tag{15}$$

4. Homogeneous Problems

Consider the homogeneous problems (π_H) and (π_H^*). It is known from the theory of homogeneous linear differential equations that all solutions $u \in C^n$ of the equation

$$L[u]=0 \tag{16}$$

form a n-dimensional function space K. The solutions $v \in C^n$ of $L^*[v]=0$ form another n-dimensional function space K^*. These two spaces K and K^* are mapped by the operator ρ into two subspaces $\mathbf{K}=\rho(K)$ and $\mathbf{K}^*=\rho(K^*)$ of R^{2n}.

[6] For self-adjoint problems Green's formula states

$$(L[u], v)=(u, L[v]) \qquad \text{for } \rho(u), \rho(v) \in \Omega$$

Usually one uses this relation to define self-adjointness of (π_N). The two definitions are equivalent as can be shown in a way similar to that used for the proof of $L^*=L$.

In these subspaces $\mathbf{K}$ and $\mathbf{K}^*$ the operator ρ has a unique inverse ρ^{-1}; hence, $K=\rho^{-1}(\mathbf{K})$ and $K^*=\rho^{-1}(\mathbf{K}^*)$. This is immediately clear because $\rho(u)=0$ implies that all initial values $u^{(\nu)}(a)$ $(\nu=0,1, \ldots, n-1)$ are zero; hence $u(t)=0$ owing to the uniqueness theorem for the initial value problem of eq (16).

The operator L and the bilinear form $\varphi(x,y)$ determine the ϕ-orthogonal space $\mathbf{K}=\rho(K)^{\perp}$ of $\mathbf{K}$.

THEOREM 1: *The subspaces $\rho(K^*)$ and $\mathbf{K}^{\perp}=\rho(K)^{\perp}$ of R^{2n} are identical, i.e.,*

$$\mathbf{K}^{\perp}=\rho(K^*). \tag{17}$$

PROOF: For any two functions $u(t) \in K$, $v(t) \in K^*$ Green's formula gives $\varphi(\rho(u), \rho(v))=0$, which implies

$$\rho(K^*) \in \rho(K)^{\perp}=\mathbf{K}^{\perp}. \tag{18}$$

Since the mapping ρ is regular in K and K^*, it follows on the other hand that

$$\dim \rho(K)=\dim \mathbf{K}=n$$

and

$$\dim \rho(K^*)=\dim \mathbf{K}^*=n.$$

Thus, we find

$$\dim \mathbf{K}^{\perp}=2n-\dim \mathbf{K}=n,$$

and the inclusion (18) must constitute equality.

To this point the boundary conditions in (π_H) and (π_H^*) have not been used. To consider them now we introduce the space A of all admissible functions of (π_H), i.e., the space of all functions $u(t) \in C^n$ with $\rho(u) \in \Omega$. Thus, the solutions of (π_H) are exactly the functions

$$u(t) \in A \cap K. \tag{19a}$$

For these functions the boundary value vectors are

$$\rho(u) \in \Omega \cap \rho(K). \tag{19b}$$

We handle (π_H^*) in the same way. Let A^* be the space of all $v(t) \in C^n$ for which $\rho(u) \in \Omega$. Then, the solutions of (π_H^*) are exactly the functions

$$v(t) \in A^* \cap K^*. \tag{20a}$$

In view of relation (17) the boundary value vectors of these functions are

$$\rho(v) \in \Omega^{\perp} \cap \mathbf{K}^{\perp}=\Omega^{\perp} \cap \rho(K)^{\perp}. \tag{20b}$$

Thus, the number of linearly independent solutions of (π_H) or (π_H^*) is given by

$$k=\dim (A \cap K)=\dim (\Omega \cap \rho(K)) \tag{21}$$

and

$$k^*=\dim (A \cap K^*)=\dim (\Omega^{\perp} \cap \rho(K)^{\perp}), \tag{22}$$

respectively.

THEOREM 2: *The numbers k and k^* of linearly independent solutions of (π_H) and (π_H^*) are connected by the relation*

$$k^* = m - n + k. \tag{23}$$

To prove this we use the following well-known lemma from linear algebra:

Let A be a linear space and A_1, A_2 two subspaces, then

$$\dim A_1 + \dim A_2 = \dim(A_1 \cap A_2) + \dim (A_1 + A_2), \tag{24}$$

where $A_1 + A_2$ is the linear closure of A_1 and A_2.

This relation is employed in a slightly different form. We introduce the orthogonal subspaces $A_1^\perp \in A$ and $A_2^\perp \in A$ of A_1 and A_2 with respect to a given (nondegenerate) bilinear form. Whence, we have

$$\dim A_1^\perp = n - \dim A_1$$

and

$$\dim (A_1^\perp \cap A_2^\perp) = \dim (A_1 + A_2)^\perp = n - \dim (A_1 + A^2).$$

Thus (24) becomes

$$\dim A_2 - \dim (A_1 \cap A_2) = \dim A_1^\perp - \dim (A_1^\perp \cap A_2^\perp). \tag{25}$$

In the present case,

$$A = R^{2n}, \qquad A_1 = \Omega, \qquad A_2 = \rho(K);$$

therefore

$$\dim A_1 = 2n - m, \qquad \dim A_1^\perp = m, \qquad \dim A_2 = n;$$

hence,

$$n - \dim (\Omega \cap \rho(K)) = m - \dim (\Omega^\perp \cap \rho(K)^\perp),$$

which completes the proof.

5. The Nonhomogeneous Problem

Consider the nonhomogeneous problem (π_N) together with its adjoint homogeneous problem (π_H^*).

We shall prove the "alternative" theorem of the theory of systems of linear equations.

THEOREM 3 (*Alternative theorem*): *(π_N) has a solution if and only if $(f,v) = 0$ for all solutions $v(t)$ of (π_H^*).* This can be expressed in the form [7]

$$L[A] = (A^* \cap K^*)^{\perp\!\!\!\perp}. \tag{26}$$

Here the symbol $\perp\!\!\!\perp$ denotes orthogonality in the function space C° with respect to the scalar product (f,g).

[7] $L[A]$ is the subspace of C° containing all functions $L[u]$ with $u \in A$.

PROOF: (1) Assume (π_N) is solvable and $u(t)$ is a solution. Then, for any solution $v(t)$ of (π_H^*), it follows from Green's formula that

$$(f,v) = (L[u],v) = (u,L^*[v]) + \varphi(\rho(u),\rho\ (v)).$$

Since $L^*[v] = 0$, then $(u,L^*[v]) = 0$; and from $\rho(u) \in \Omega$ and $\rho(v) \in \Omega^\perp$ it follows that $\varphi(\rho(u),\ \rho(v)) = 0$. Thus it is seen that $(f,v) = 0$.

(2) Assume that $(f,v) = 0$ for all solutions of (π_H^*). We shall show that (π_N) has a solution.

From the existence theorem for ordinary differential equations it is known that the differential equation $L[u] = f$ must have at least one solution $u_0(t)$. For this function $u_0(t)$ and any solution $v(t)$ of (π_H^*) we find from Green's formula that

$$\begin{aligned} 0 = (f,\ v) = (L[u_0],\ v) &= (u_0,\ L^*[v]) + \varphi(\rho(u_0),\ \rho(v)) \\ &= \varphi(\rho(u_0),\ \rho(v)). \end{aligned}$$

Thus $\rho(u_0)$ is φ-orthogonal to all $\rho(v) \in \Omega^\perp \cap \rho(K)^\perp$ and

$$\rho(u_0) \epsilon (\Omega^\perp \cap \rho(K)^\perp) = \Omega + \rho(K)$$

where the $+$ sign again denotes the linear closure of Ω and $\rho(K)$.

This inclusion states the existence of two vectors $\mathbf{u}_1 \in \Omega$, $\mathbf{u}_2 \in \rho(K)$ such that $\rho(u_0) = \mathbf{u}_1 + \mathbf{u}_2$. From the fact that ρ has a unique inverse ρ^{-1} in $\mathbf{K}$ it follows that there is exactly one function $u_2(t) = \rho^{-1}(\mathbf{u}_2) \in K$ for which $L[u_2] = 0$ and $\rho(u_2) = \mathbf{u}_2$. We form a new function $u(t) = u_0(t) - u_2(t)$. This function $u(t)$ is a solution of (π_N), for obviously $L[u] = f$ and $\rho(u) = \rho(u_0) - \rho(u_2) = \mathbf{u}_1 \in \Omega$.

Taking the orthogonal complement on both sides of (26) we find

$$L[A]^{\perp\!\!\!\perp} = (A^* \cap K^*)^{\perp\!\!\!\perp\ \perp\!\!\!\perp} = A^* \cap K^*. \tag{27}$$

Thus we get the following

COROLLARY: *If for a given function $v(t) \in C^\circ$,*

$$(v(t),L[u]) = 0 \quad \text{for every} \quad u(t) \in A,$$

then $L^*[v] = 0$, $\rho(v) \in \Omega$ and in particular $v(t) \in C^n$.

6. Green's Function

The solution of the nonhomogeneous problem (π_N)—provided that it is solvable—is determined up to a solution of the homogeneous problem (π_H). In order to obtain an explicit formula for the solution of (π_N), it is necessary to assume an additional condition which uniquely determines the solution of (π_N). Furthermore, it is desirable to modify (π_N) in such a way that it is always solvable.

With these two aims in mind we introduce an orthonormalized basis in each of the spaces

$$A \cap K \qquad \text{and} \qquad A^* \cap K^*.$$

Let

$$u_1, u_2, \ldots, u_k \quad \text{and} \quad v_1, v_2, \ldots, v_k \tag{28}$$

be these two bases. With the functions $v_\nu(t)$ we define the new boundary value problem

$$L[u]=f(t)-\sum_{\nu=1}^{k^*}(f(t),v_\nu(t))v_\nu(t),\ \rho(u)\in\Omega. \tag{29}$$

Owing to the orthonormality of the $v_\nu(t)$, the right-hand side of the differential equation is orthogonal to all $v_\nu(t)$ $(\nu=1, \ldots, k)$, and thus to all solutions of (π_H^*). Therefore (29) is solvable for every function $f(t)\in C^0$. To find an additional condition which uniquely determines the solution of problem (29), we recall the following fact: If $u(t)$ is a solution of (29), then

$$u(t)+\sum_{\mu=1}^{k}\gamma_\mu u_\mu(t)$$

is also a solution for any real coefficients γ_μ. We can determine the γ_μ in such a way that

$$\left(u(t)+\sum_{\mu=1}^{k}\gamma_\mu u_\mu(t),u_\nu(t)\right)=0 \qquad (\nu=1, \ldots, k).$$

Owing to the orthonormality of the u_μ, these relations are satisfied for

$$\gamma_\mu=-(u(t),u_\mu(t)) \qquad (\mu=1, \ldots, k).$$

This shows that there is always a solution of (29), which is orthogonal to all u_μ and thus to the whole space $A\cap K$; obviously, this solution is uniquely determined.

Consider instead of (π_N) the following problem:

$$L[u]=f(t)-\sum_{\mu=1}^{k^*}(f,v_\mu)\,v_\mu(t)$$

$$\rho(u)\in\Omega \tag{$\tilde{\pi}_N$}$$

$$(u(t),u_\mu(t))=0 \qquad (\mu=1, \ldots, k).$$

We know that this problem $(\tilde{\pi}_N)$ has exactly one solution $u(t)\in C^n$ for every function $f(t)\in C^0$. The correspondence between $u(t)$ and $f(t)$ is denoted by an operator G, i.e.,

$$u(t)=G[f(t)]. \tag{30}$$

This operator is obviously linear. It will be shown that G is an integral operator of the form

$$u(t)=G[f(t)]=\int_a^b g(t,\tau)\,f(\tau)\,d\tau, \tag{31}$$

where $g(t,\tau)$ is the so-called Green's function in the generalized sense.

To prove this, we consider functions $g(t,\tau)$ having the following properties:

(P 1) $g(t,\tau)$ is a continuous function of t, τ for $a\leqq t,\tau\leqq b$;

(P 2) For every arbitrary but fixed $\tau=\tau_0, a\leqq\tau_0\leqq b$, $g(t,\tau_0)$ possesses n continuous derivatives for $a\leqq t\leqq\tau_0$ and $\tau_0\leqq t\leqq b$ such that

$$\frac{\partial^\nu g(\tau_0+0,\tau_0)}{\partial t^\nu}-\frac{\partial^\nu g(\tau_0-0,\tau_0)}{\partial t_0}=0 \qquad (\nu=0,1,\ldots,n-2)$$

$$\frac{\partial^{n-1} g(\tau_0+0,\tau_0)}{\partial t^{n-1}}-\frac{\partial^{n-1} g(\tau_0-0,\tau_0)}{\partial t^{n-1}}=\frac{1}{a_0(\tau_0)};$$

(P 3)[8] For every arbitrary but fixed $\tau=\tau_0$ in $a\leqq\tau_0\leqq b$,

$$L_t[g(t,\tau_0)]=-\sum_{\mu=1}^{k^*}v_\mu(t)\,v_\mu(\tau_0)$$

for $a\leqq t\leqq b$ but $t\neq\tau_0$;

(P 4)[8] For every arbitrary but fixed $\tau=\tau_0$ in $a<\tau_0<b$,

$$\rho_t(g(t,\rho_0))\in\Omega;$$

(P 5) For all functions $u_\mu(t)$ $(\mu=1, \ldots, k)$ of the basis (28)

$$\int_a^b g(t,\tau)\,u_\mu(t)\,dt\equiv 0$$

identically in $a\leqq\tau\leqq b$.

It shall be proved that there exists exactly one function $g(t,\tau)$ having these properties which satisfies (31).

THEOREM 4: *If there exists a function $g(t,\tau)$ possessing properties (P 1) through (P 5), it is uniquely determined.*

PROOF: Let $g_1(t,\tau)$ and $g_2(t,\tau)$ be two functions which both possess properties (P 1) through (P 5). For an arbitrary but fixed $\tau=\tau_0$ in $a<\tau<b$ we define the function

$$g_0(t)=g_1(t,\tau_0)-g_2(t,\tau_0).$$

Then, according to (P 3) it is seen that

$$L[g_0(t)]=0 \qquad \text{for} \qquad a\leqq t\leqq b, \quad t\neq\tau_0. \tag{32}$$

In view of (P 2), $g_0(t)\in C^{n-1}$; hence it follows from (32) that the jump in the $(n-1)$-th derivative of $g_0(t)$ at $t=\tau_0$ must be zero. Thus, $g_0(t)\in C^{n-1}$ and in view of (32), $g_0(t)\in K$. From (P 4) it follows that $g_0(t)\in A$; therefore, $g_0(t)\in K\cap A$. On the other hand, it can be concluded from (P 5) that $g_0(t)\in(K\cap A)^{\perp\perp}$. This is possible only if $g_0(t)\equiv 0$ for $a\leqq t\leqq b$. Because τ_0 was arbitrarily chosen in the open interval $a<\tau<b$, this is equivalent to the statement $g_1(t,\tau)=g_2(t,\tau)$ for $a\leqq t\leqq b$, $a<\tau<b$, and in view of (P 1) for the whole $a\leqq t,\tau\leqq b$.

[8] The subscripts t in $L_t[g]$ and $\rho_t(g)$ denote that the differential operator L and the mapping ρ operate on $g(t,\tau)$ regarded as a function of t.

To state the existence of $g(t, \tau)$ we first prove the following:

LEMMA 1: *There exists at least one function $h(t, \tau)$ which possesses properties (P 1), (P 2), (P 3), and (P 5).*

PROOF: For fixed arbitrary τ in $a<\tau<b$, consider the ordinary differential equation

$$L_t[g(t, \tau)]=-\sum_{\mu=1}^{k^*} v_\mu(t)v_\mu(\tau) \tag{33}$$

and the two sets of initial values

$$\left.\frac{\partial^\nu g(t, \tau)}{\partial t^\nu}\right|_{t=\tau}=0 \qquad (\nu=0, 1, \ldots, n-1) \tag{34a}$$

and

$$\left.\frac{\partial^{n-1} g(t,\tau)}{\partial t^{n-1}}\right|_{t=\sigma}=\frac{1}{a_0(\tau)}. \tag{34b}$$

Both initial value problems (33)/(34a) and (33)/(34b) possess unique solutions $h_1(t, \tau)$ and $h_2(t, \tau)$, respectively. These solutions h_1 and h_2 are continuous functions of t, τ in the whole square, $a\leqq t,\tau\leqq b$. Furthermore, they possess n continuous derivatives with respect to t in this square. Accordingly, the function

$$h_3(t,\tau)=\begin{cases} h_1(t,\tau) & \text{for} \quad a\leqq t\leqq\tau \\ h_2(t,\tau) & \text{for} \quad \tau\leqq t\leqq b \end{cases}$$

obviously possesses properties (P 1), (P 2), and (P 3).

We define the coefficient-functions

$$c_\mu(\tau)=(h_3(t, \tau), u_\mu(t)) \qquad (\mu=1, \ldots, k),$$

which in view of the continuity properties of $h_3(t, \tau)$ are continuous functions of τ. Hence, the function

$$h(t,\tau)=h_3(t,\tau)-\sum_{\mu=1}^{k} c_\mu(\tau)u_\mu(t)$$

evidently satisfies properties (P 1), (P 2), and (P 3). Furthermore (P 5) obtains, since

$$(h(t, \tau), u_\mu(t))=(h_3(t, \tau), u_\mu(t))-c_\mu(\tau)=0 \qquad (\mu=1, \ldots, k),$$

owing to the orthonormality of the $u_\mu(t)$.

LEMMA 2: *There exists a linear transformation which maps every vector $\mathbf{x}\in\Omega+\rho(K)$ into one vector $\mathbf{y}\in\rho(K)$ such that*

$$(\rho^{-1}(\mathbf{y}), u_\mu(t))=0 \qquad (\mu=1, \ldots, k). \tag{35}$$

PROOF: Let $\mathbf{x}\in\Omega+\rho(K)$ be given. Then there exist at least two vectors $\mathbf{x}_1,\mathbf{x}_2$ for which

$$\mathbf{x}=\mathbf{x}_1+\mathbf{x}_2 \qquad \mathbf{x}_1\in\Omega \qquad \mathbf{x}_2\in\rho(K).$$

$\mathbf{x}_1$ and $\mathbf{x}_2$ are determined to within a vector belonging to the intersection $\Omega\cup\rho(K)$. With coefficients

$$c_\mu=(\rho^{-1}(\mathbf{x}_2), u_\mu(t)) \qquad (\mu=1, \ldots, k),$$

the following new vectors are defined:

$$\mathbf{y}_1=\mathbf{x}_1+\sum_{\mu=1}^{k} c_\mu\rho(u_\mu), \quad \mathbf{y}_2=\mathbf{x}_2-\sum_{\mu=1}^{k} c_\mu\rho(u_\mu).$$

Since

$$\sum_{\mu=1}^{k} c_\mu\rho(u_\mu)\in\Omega\cap\rho(K),$$

then

$$\mathbf{x}=\mathbf{y}_1+\mathbf{y}_2 \qquad \mathbf{y}_1\in\Omega \qquad \mathbf{y}_2\in\rho(K), \tag{36}$$

and, furthermore, in view of the orthonormality of the u_μ,

$$\begin{aligned}(\rho^{-1}(\mathbf{y}_2), u_\mu)&=(\rho^{-1}(\mathbf{x}_2), u_\mu)-\sum_{\nu=1}^{k} c_\nu(\rho^{-1}\rho(u_\nu), u_\nu)\\ &=(\rho^{-1}(\mathbf{x}_2), u_\mu)-c_\mu=0 \qquad (\mu=1, \ldots, k).\end{aligned} \tag{37}$$

The two vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ are uniquely determined by conditions (36) and (37). This is rather obvious; for any other decomposition $\mathbf{x}=\mathbf{y}_1'+\mathbf{y}_2'$ of $\mathbf{x}$ which satisfies (36) and (37) it follows that $\rho^{-1}(\mathbf{y}_2-\mathbf{y}_2')\in A\cap K$. However, the relations

$$(\rho^{-1}(\mathbf{y}_2-\mathbf{y}_2'), u_\mu)=0 \qquad (\mu=1, \ldots, k)$$

obtain only if $\mathbf{y}_2=\mathbf{y}_2'$ which then implies $\mathbf{y}_1=\mathbf{y}_1'$.

Hence, it has been shown that we can associate with every $\mathbf{x}\in\Omega+\rho(K)$ exactly one vector $\mathbf{y}_2\in\rho(K)$ which fulfills (36) and (37). This was the statement of the lemma.

It is necessary at this point to obtain an extension of Green's formula for functions having discontinuity properties similar to (P 2). Let the function $y(t)$ possess the following properties:

a) $y(t)$ possesses n continuous derivatives for $a\leqq t\leqq\tau$, $\tau\leqq t\leqq b$, where τ is a certain fixed point.

b) $y^{(\nu)}(\tau+0)-y^{(\nu)}(\tau-0)=0 \qquad (\nu=0,1, \ldots,n-2),$

$$y^{(n-1)}(\tau+0)-y^{(n-1)}(\tau-0)=c,$$

where c is a certain fixed constant.

For any arbitrary chosen $v(t)\in C^n$, it can be shown by repeated partial integration that

$$(L[y],v)-(y,L^*[v])=\varphi(\rho(y), \rho(v))-c\ a_0(\tau)v(\tau). \tag{38}$$

This is the generalized Green's formula.

THEOREM 5: *There exists a function $g(t,\tau)$ which possesses the five properties* (P 1) *through* (P 5).

PROOF: Let $h(t,\tau)$ be the function which has properties (P 1), (P 2), P 3), and (P 5). For every arbitrary but fixed τ in $a<\tau<b$, insert $h(t,\tau)$ and the basic functions $v_\nu(t)$ $(\nu=1, \ldots, k)$ in the generalized Green's formula:

$$\begin{aligned}(L_t[h], v_\nu)&=(h, L^*[v_\nu])+\varphi(\rho_t(h),\rho(v_\nu))-v_\nu(\tau)\\ &=\varphi(\rho_t(h), \rho(v_\nu))-v_\nu(\tau) \qquad (\tau=1, \ldots, k^*).\end{aligned}$$

On the other hand, for $a \leqq t \leqq b$, $t \neq \tau$, according to (P 3),

$$(L_t[h], v_\nu) = -\sum_{\mu=1}^{k^*} (v_\mu(t), v_\nu(t)) v_\nu(\tau) = -v_\nu(\tau),$$

and in view of the orthonormality of the $v_\nu(t)$

$$\varphi(\rho_t(h), \rho(v_\nu)) = 0 \qquad (\nu = 1, \ldots, k^*).$$

This is equivalent to

$$\rho_t(h) \in (\Omega^\perp \cup \rho(K)^\perp)^\perp = \Omega + \rho(K).$$

We now consider the (continuous) linear transformation from $\Omega + \rho(K)$ into $\rho(K)$ which exists by lemma 2. Let $\mathbf{h} \in \rho(K)$ with

$$(\rho^{-1}(\mathbf{h}), u_\mu) = 0 \qquad (\mu = 1, \ldots, k) \tag{39}$$

be the image vector of $\rho_t(h)$ under this transformation. Then $\mathbf{h}$ is a continuous function of τ for $a \leqq \tau \leqq b$. This is obvious because $\rho_t(h)$ depends continuously on τ, and the correspondence between $\rho_t(h)$ and $\mathbf{h}$ is continuous. Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be a basis of $\rho(K)$ and

$$\mathbf{h} = \sum_{\mu=1}^{n} d_\mu(\tau) \mathbf{x}_\mu.$$

It is clear that the coefficients $d_\mu(\tau)$ are continuous functions of τ for $a \leqq \tau \leqq b$. Hence, the (uniquely determined) function

$$\hat{h}(t, \tau) = \rho^{-1}(\mathbf{h}) = \sum_{\mu=1}^{n} d_\mu(\tau), \qquad \rho^{-1}(\mathbf{x}_\mu) \in K$$

is continuous in both variables t and τ for $a \leqq t, \tau \leqq b$.

We define the function

$$g(t,\tau) = h(t,\tau) - \hat{h}(t,\tau).$$

Then $g(t,\tau)$ obviously possesses properties (P 1), (P 2), and (P 3). (P 4) follows from

$$\rho_t(g) = \rho_t(h) - \mathbf{h} \in \Omega.$$

Finally, as a consequence of (39),

$$(g(t,\tau), u_\mu(t)) = (h(t,\tau), u_\mu(t)) - (\hat{h}(t,\tau), u_\mu(t)) = 0$$

for $\mu = 1, \ldots, k$, since $h(t,\tau)$ fulfills (P 5).

Thus $g(t,\tau)$ possesses properties (P1) through (P 5). This proves the theorem.

Now we are in a position to prove the following final theorem:

THEOREM 6: *Let $g(t,\tau)$ be the uniquely determined function with the properties* (P1) *through* (P5); *then*

$$u(t) = G[f(t)] = \int_a^b g(t, \tau) f(\tau) d\tau,$$

for every given $f(\tau) \in C^\circ$, constitutes the (uniquely determined) solution of $(\hat{\pi}_N)$.

PROOF: Let $v(t) \in A^*$ be an arbitrary function. For any fixed τ in $a < \tau < b$ we apply the generalized Green's formula (38) to $g(t,\tau)$ and $v(t)$. With regard to (P 4) and $\rho(v) \in \Omega^\perp$ it is seen that

$$(L_t[g], v) = (g, L^*[v]) - v(\tau);$$

thus, in view of (P3),

$$(L_t[g], v) = -\sum_{\nu=1}^{k^*} (v, v_\nu) v_\nu(\tau) = (g, L^*[v]) - v(\tau)$$

for $a \leqq t \leqq b$, $a < \tau < b$, $t = \tau$. We multiply this equation by a continuous function $f(\tau)$ and integrate over τ. Because g is a continuous function of t and τ, it follows that

$$-\sum_{\nu=1}^{k^*} (v, v_\nu)(v_\nu, f) = (G[f], L^*[v]) - (f, v). \tag{40}$$

For $f(t)$ we consider the boundary value problem (29) which is solvable no matter how $f(t)$ was chosen. Let $u_1(t)$ be a solution. For the given function $v(t) \in A^*$ we obtain

$$(L[u_1], v) = (f, v) - \sum_{\nu=1}^{k^*} (f, v_\nu)(v_\nu, v),$$

and thus together with (40)

$$(L[u_1], v) = (G[f], L^*[v]).$$

Applying Green's formula (11) to the left-hand side we conclude that

$$(L^*[v], u_1) = (G[f], L^*[v]) - \varphi(\rho(u_1), \rho(v)),$$

and because $\varphi(\rho(u_1), \rho(v)) = 0$, then

$$(G[f] - u_1, L^*[v]) = 0.$$

This relation is valid for every function $v(t) \in A^*$. Hence, it follows from the corollary to the alternative theorem that $G[f] - u_1 \in C^n$; thus, $u(t) = G[f] \in C^n$. Furthermore, it follows that

$$L[G[f] - u_1] = 0, \qquad \rho(G[f] - u_1) \in \Omega,$$

which is equivalent to

$$L[G[f]] = L[u_1] = f - \sum_{\nu=1}^{k^*} (f, v_\nu) v_\nu,$$
$$\rho(G[f]) \in \Omega.$$

Finally, it follows from (P5) and (P1) that

$$(u(t), u_\mu(t)) = 0 \qquad (\mu = 1, \ldots, k).$$

In applying theorem 6 to the adjoint problem $(\hat{\pi}_N{}^*)$ we confirm that for any $f_1(t) \in C^\circ$ the uniquely

determined solution $v(t)$ of $(\hat{\pi}_N^*)$ is given by

$$v(t)=G^*[f_1]=\int_a^b g^*(t,\tau)f^1(\tau)d\tau. \tag{41}$$

We will show that the integral operator G^* is the adjoint operator of G, i.e.,

$$(G[f],f_1)=(f,\ G^*[f_1]) \tag{42}$$

for all $f_1, f\in C^\circ$.

Let $v(t)=G^*[f_1]$, i.e.,

$$L^*[G^*[f_1]]=f_1-\sum_{\nu=1}^{k}(f_1,u_\nu)u_\nu$$

$$\rho(G^*[f_1])\in\Omega^\perp;\qquad (G^*[f_1],v_\mu)=0\quad(\mu=1,\ \ldots,\ k^*).$$

WASHINGTON, D.C. (Paper 64B2–24)

With another continuous function $f(t)$ we find from Green's formula (11),

$$\begin{aligned}(G[f],f_1)&=(G[f],L^*[G^*[f_1]])+\sum_{\nu=1}^{k}(f_1,u_\nu)(G[f],u_\nu)\\&=(L[G[f]],G^*[f_1])+\varphi(\rho(G[f]),\rho(G^*[f_1]))\\&\quad+\sum_{\nu=1}^{k}(f_1,u_\nu)(G[f],u_\nu)\\&=(f,G^*[f_1])+\sum_{\mu=1}^{k^*}(f,v_\mu)(v_\mu,G^*[f_1])\\&\quad+\varphi(\rho(G[f]),\rho(G^*[f_1]))+\sum_{\nu=1}^{k}(f_1,u_\nu)(G[f],u_\nu).\end{aligned}$$

Since $G[f_1]$ and $G^*[f_1]$ are solutions of $(\hat{\pi}_N)$ and $(\hat{\pi}_N^*)$, respectively, the last three quantities must vanish, which leads to (42).

From (42) follows for the corresponding Green's functions,

$$g^*(t,\tau)=g(\tau,t).$$

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 64B, No. 2, April-June 1960

Criteria for the Existence and Equioscillation of Best Tchebycheff Approximations

John R. Rice [1]

(January 19, 1960)

Let $F(a,x)=\Sigma a_i\varphi_i(x)$ where $\varphi_i(x)$ are continuous in [0,1] and the a_i are real parameters. The following theorem answers the principal questions of a general nature in the theory of Tchebycheff approximations: THEOREM. Let $\{\varphi_i(x)\}$ form a Tchebycheff set and let $f(x)$ be an arbitrary function continuous on [0,1]. Then (A) $f(x)$ possesses a best approximation, (B) a necessary and sufficient condition that $F(a^*,x)$ be a best approximation to $f(x)$ is that $\max|F(a^*,x)-f(x)|$ alternates at least n times in [0,1], (C) the best approximation to $f(x)$ is unique. A. Haar [Math. Ann. **78**, 43–56 (1928)] posed and answered the following question: What conditions on F are necessary and sufficient for theorem C to be valid? The condition he found is that $\{\varphi_i(x)\}$ must form a Tchebycheff set. This paper poses and answers the following three questions: (1) What conditions on F are necessary and sufficient for theorem A to be valid? (2) What conditions on F are necessary and sufficient for theorem B to be valid? (3) What conditions on F are necessary and sufficient for both theorems A and B to be valid?

This paper does not tacitly assume that the a_i may assume all values.

1. Let T be a set of n functions $\varphi_1(x)$, $\varphi_2(x)$, . . ., $\varphi_n(x)$ continuous on [0,1] and let $F(a,x)=\Sigma_{i=1}^{n}a_i\varphi_i(x)$ be called a *T-polynomial*. $F(a,x)$ is said to be a *nontrivial T-polynomial* if $\Sigma_{i=1}^{n}|a_i|>0$. It is assumed that there is no nontrivial T-polynomial which is identically zero. The *parameter set* P of F is the domain of the a_i. T is said to be a *Tchebycheff set* if every nontrivial T-polynomial has at most $n-1$ zeros. $\max|F(a,x)-f(x)|$ is said to *alternate n times* if there are $n+1$ points $0\leq x_1<x_2< \ . \ . \ . \ <x_{n+1}\leq 1$ such that

$$F(a,x_j)-f(x_j)=-[F(a,x_{j+1})-f(x_{j+1})]$$
$$=\pm\max|F(a,x)-f(x)|.$$

All maximums and minimums are taken over $x\in[0,1]$ unless otherwise stated, and all summations are from 1 to n unless otherwise stated. Braces, $\{\ \}$, denote sets and sequences, and $\{x|\ .\ .\ .\ \}$ is read as "the set of x such that . . .".

The following theorem answers the principal questions of a general nature in the theory of Tchebycheff approximations.

THEOREM: *Let T be a set of n functions and let $f(x)$ be an arbitrary function continuous on [0,1]. Then*

A. $f(x)$ possesses a best approximation,

B. a necessary and sufficient condition that $F(a^*,x)$ be a best approximation to $f(x)$ is that $\max|F(a^*,x)-f(x)|$ alternates at least n times.

C. the best approximation to $f(x)$ is unique.

A. Haar [2] has posed and answered the following question: What conditions on F are necessary and sufficient in order that theorem C be valid for every continuous $f(x)$? The condition he found is that T must be a Tchebycheff set.

Three similar questions are posed in this paper. They are:

(1) What conditions on F are necessary and sufficient for theorem A to be valid for every continuous $f(x)$?

(2) What conditions on F are necessary and sufficient for both theorems A and B to be valid for every continuous $f(x)$?

(3) What conditions on F are necessary and sufficient for theorem B to be valid for every continuous $f(x)$?

These questions are answered by theorems 1, 2, and 3, respectively.

In this study it has not been assumed that $P=E_n$, and indeed the topological nature of P plays a central role in the analysis.

2. The following definition will be useful:

DEFINITION: *F has property Z if $a\neq a^*$ implies that $F(a,x)-F(a^*,x)$ has at most $n-1$ zeros in [0,1].*

Note that this does not imply that T is a Tchebycheff set for the parameters are restricted to P.

The first theorem gives an answer to the first question posed.

THEOREM 1: *Theorem A is valid if and only if P is closed.*

PROOF: Let $d=\inf_{a\in P}\max|F(a,x)-f(x)|$, then it can be shown [3] that there is an $M<\infty$ such that $\Sigma|a_i|>M$ implies $\max|F(a,x)-f(x)|>d+1$. $f(x)$ has a best approximation if there is an $a^*\in P$ such that

$$\max|F(a^*,x)-f(x)|=\inf\max|F(a,x)-f(x)|.$$

Let

$$P_M=\{a|a\in P,\sum|a_i|\leq M\}.$$

[1] Part of this work was done while the author was an NRC–NBS Postdoctoral Research Associate.

[2] A. Haar, Die Minkowskische Geometrie und die Annäherung an stetige Functionnen, Math. Ann. **78**, 43–56 (1928).

[3] N. I. Achieser, Theory of approximation, p. 10 (Ungar Publishing Co., New York, N.Y., 1956).

Clearly a may be restricted to values from P_M. If P is closed, then $\inf_{a \in P_M} \max|F(a,x)-f(x)|$ is attained and $f(x)$ possesses a best approximation.

Assume that P is not closed and let a' be a boundary point of P not contained in P. Consider the function $\Sigma a_i' \varphi_i(x)$. If theorem A is valid, then it has a best approximation, say $F(a^*,x)$. Now it is clear that

$$\inf_{a \in P} \max |F(a,x)-\sum a_i' \varphi_i(x)|=0,$$

hence

$$\sum (a_i'-a_i^*)\varphi_i(x)=0 \text{ for } x \in [0,1].$$

This possibility is precluded by one of the initial assumptions; so a contradiction has been reached.

The following example shows that the assumption that no linear combination of the $\varphi_i(x)$ be identically zero is necessary for the validity of theorem 1. Let $\varphi_1(x)=1$, $\varphi_2(x)=1$, and let P be the subset of E_2 consisting of $\{(x,0)|x \neq 1\}$ and $(0,1)$. Every continuous function has a unique best approximation, but P is not closed.

3. In order to answer the second question, three lemmas will be established. The first is

LEMMA 1: *If theorem B is valid, then F has property Z.*

PROOF: Assume $F(a,x)-F(a^*,x)$ has n zeros. Let

$$M(x)=\tfrac{1}{2}[F(a,x)+F(a^*,x)];$$

$$d=\tfrac{1}{2} \max |F(a,x)-F(a^*,x)|,$$

and let $0 \leq x_1 < x_2 < , \ldots, < x_{n+1} \leq 1$ be n points where $F(a,x)-F(a^*,x)=0$, along with one point where $F(a,x)-F(a^*,x) \neq 0$, say $F(a^*,x_2)>F(a,x_2)$ for concreteness. Set

$$\delta=\tfrac{1}{4} \min (x_1, x_2-x_1, \ldots, x_{n+1}-x_n, 1-x_{n+1}),$$

with x_1 or $1-x_{n+1}$ omitted if they are zero.

A function $f(x)$ continuous on $[0,1]$ will be defined as follows:

$$f(x)=M(x) \quad \text{in the intervals} \quad [x_j+\delta, \quad x_{j+1}-\delta],$$

$$j=1,2, \ldots, n;$$

$$f(x)=M(x) \quad \text{in} \quad [0,x_1-\delta] \quad \text{if} \quad x_1>0;$$

$$f(x)=M(x) \quad \text{in} \quad [x_{n+1}+\delta, 1] \quad \text{if} \quad x_{n+1}<1;$$

in the remaining intervals $f(x)$ satisfies

$$f(x_j)=F(a^*,x_j)+(-1)^{j+1}(3d/2)$$

and

$$|F(a,x)-f(x)|<3d/2, \quad |F(a^*,x)-f(x)|<3d/2 \text{ for } x \neq x_j.$$

Now $\max |F(a^*,x)-f(x)|$ alternates exactly n times and $\max |F(a,x)-f(x)|$ alternates exactly $n-2$ times. By theorem B, $F(a^*,x)$ is a best approximation to $f(x)$, and since

$$\max |F(a^*,x)-f(x)|=\max |F(a,x)-f(x)|,$$

so is $F(a,x)$. This contradicts the validity of theorem B.

For the purposes of the next lemma an *axial set* will be defined. A set in E_n is said to be axial if every neighborhood of each point a of the set contains a point of the set on each half-axis (positive and negative) of the E_n coordinate system with origin translated to a.

Since the validity of theorem B implies that P has property Z, the parameter space P may be identified with the function values of F at n fixed distinct points in $[0,1]$. Let $0<x_1<x_2<, \ldots, <x_n<1$ be n such points and let P' denote the set identified with P.

LEMMA 2: *If theorem B is valid and P' is closed, then P' is an axial set.*

PROOF: Let $\epsilon>0$ and a point $b^* \in P'$ be given. $F(b^*,x)$ shall denote $F(a^*,x)$ where $a^* \in P$ is identified with b^*. Let

$$d=\min [x_1, x_2-x_1, \ldots, x_n-x_{n+1}, 1-x_n].$$

A sequence of functions $\{f_m(x)|m=2,3, \ldots\}$, each continuous on $[0,1]$, will be defined as follows:

$$f_m(x)=F(b^*,x)+(-1)^j\epsilon/3 \quad \text{for} \quad x \in [0,x_1-d/m],$$

$$x \in [x_j+d/m, x_{j+1}-d/m], \quad j=1,2, \ldots, k-2;$$

$$f_m(x)=F(b^*,x)+(-1)^{k-1}\epsilon/3 \quad \text{for}$$

$$x \in [x_{k-1}+d/m, x_{k+1}-d/m];$$

$$f_m(x)=F(b^*,x)+(-1)^{j+1}\epsilon/3 \text{ for } x \in [x_n+d/m,1],$$

$$x \in [x_j+d/m, x_{j+1}-d/m], \quad j=k+1, \ldots, n;$$

$F(x)$ is defined in the remaining intervals so that $|f_m(x)-F(b^*,x)|<\epsilon/3$ and $f_m(x)$ is continuous in $[0,1]$.

Since P' is closed, P is closed and every function has a best approximation by lemma 1 and theorem 1. Let $F(b_m,x)$ be the best approximation to $f_m(x)$. Now it is clear that $\operatorname{sgn}[F(b_m,x_k)-F(b^*,x_k)]=(-1)^{k-1}$ and that $F(b_m,x)-F(b^*,x)$ has a zero in

$$[x_j-d/m, x_j+d/m], j=1,2, \ldots, k-1, k+1, \ldots, n.$$

Since P' is closed (and clearly the b_m are bounded), $\{b_m\}$ contains a convergent subsequence with a limit in P'. Let b_0 be this limit. Now $F(b^*,x_j)-F(b_0,x_j)=0$, $j=1,2, \ldots, k-1,k+1, \ldots, n$. By theorem B, $\max |F(b_m,x)-f_m(x)|$ alternates n times; hence

$$\max|F(b_m,x)-F(b^*,x)|>\epsilon/3,$$

since $\max |F(b^*,x)-F_m(x)|$ alternates only $n-1$ times. Therefore,

$$\max |F(b_0,x)-F(b^*,x)| \geq \epsilon/3 \text{ and } b_0 \neq b^*.$$

By actual construction it has been shown that there is a $b_0 \in P'$ such that b_0 is on the kth coordinate axis (positive or negative according to whether k is odd or even) at a distance less than ϵ for any $\epsilon>0$. Clearly this construction may be modified to obtain a point in P' on the kth axis with opposite sign. Hence P' is axial.

LEMMA 3: *Let P' be a closed axial set; then $P'=E_n$.*

PROOF: The proof is based on the following assertion: Every closed n-sphere whose boundary passes through the origin in E_{n+1} contains a segment of a coordinate axis which is attached to the origin. That this assertion is true is seen as follows. Consider the tangent plane of the sphere at the origin. If the positive kth axis is on one side of the plane, then the negative kth axis is on the other side; i.e., $a_k x_k > 0$ implies $a_k(-x_k) < 0$. Now the tangent plane cannot contain every axis; hence there is an axis intersecting the tangent plane at the origin. Clearly this axis must intersect the sphere also, which establishes the assertion.

If P' is closed, the complement of P' is open. Let c be a point in the complement of P', and let S be the smallest closed sphere with center at c which intersects P', and let b be a point of intersection. Now S contains a segment of some coordinate axis of a coordinate system with origin at b. Since P' is axial, that segment contains a point of P' which contradicts the definition of S and hence the existence of c.

With these three lemmas the second question may be answered by the following theorem.

THEOREM 2: *Theorems A and B are both valid for F if and only if $P=E_n$ and T is a Tchebycheff set.*

PROOF: It is a classical result that theorems A and B are valid if $P=E_n$ and T is a Tchebycheff set.

From lemma 1 it is seen that the validity of theorem B implies that F has property Z. If F has property Z and theorem A is valid, then theorem 1 implies that P is closed. Now there is a one-to-one mapping between P and P', the set defined for lemma 2, which is continuous both ways, since F has property Z. Hence, if P is closed, then so is P'. From lemma 2 it follows that P' is an axial set, and lemma 3 implies that $P'=E_n$. Hence $P=E_n$ and T is a Tchebycheff set.

4. The answer to the third question is given by theorem 3. The proof of this theorem is rather complicated and since it is a special case of a theorem to be published elsewhere, the proof will not be given here. The result is given here for the sake of completeness.

THEOREM 3: *Theorem B is valid for F if and only if (i) F has property Z;*

(ii) given $a^* \in P$, $k < n$, $\{x_j | 0=x_0<x_1 < \ldots < x_{k+1}=1\}$ and ϵ with $0<\epsilon<\frac{1}{2}$ min $(x_{j+1}-x_j)$, $j=0, 1, \ldots, k$, then (a) there are $a_1, a_2 \in P$ such that, for $x \in [0,1]$, $F(a^*, x)-\epsilon < F(a_1,x) < F(a^*,x) < F(a_2,x) < F(a^*,x)+\epsilon$, (b) there are $a_3, a_4 \in P$ such that $|F(a_3,x)-F(a^*,x)| < \epsilon$, $|F(a_4,x)-F(a^*,x)| < \epsilon$, for $x \in [0,1]$ and $F(a_3,x)-F(a^*,x)$, $F(a_4,x)-F(a^*,x)$ change sign from $x_j-\epsilon$ to $x_j+\epsilon$, $j=1, 2, \ldots, k$ and have no zeros outside $[x_j-\epsilon, x_j+\epsilon]$. Further $F(a_3,0) > F(a^*,0) > F(a_4,0)$.

WASHINGTON, D.C. (Paper 64B2–25)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 64B, No. 2, April–June 1960

Note on the Solution of Riccati's Differential Equation

H. Herbert Howe

(February 4, 1960)

Three recurrence formulas are developed, giving the solution of a particular case of Riccati's equation in power series valid in the neighborhoods of 0, ∞, and an arbitrary point, respectively. The first two were programed for computation on the SEAC.

1. Introduction

A particular case of Riccati's differential equation is important in the theory of radio propagation. In the notation often used in that subject, omitting subscripts irrelevant in the present paper, the equation is

$$\frac{d\delta}{d\tau}-2\delta^2\tau+1=0 \tag{1}$$

where δ will be considered the independent variable, and both variables are complex. The equation may be written

$$\frac{d\tau}{d\delta}=\frac{1}{2\delta^2\tau-1} \tag{2}$$

Since this is a first-order differential equation, there should be one constant of integration. If the value of τ is arbitrarily chosen for some value of δ, the differential equation ought to define τ for all other values of δ. We may, for example, arbitrarily select $\tau(0)$ or $\tau(\infty)$, and from it derive τ for all other points.

If there exists in the complex δ-plane any point P such that as we approach it the value of $\delta^2\tau$ approaches $\frac{1}{2}$, then the derivative $d\tau/d\delta$ approaches infinity, and P is a singular point of τ. Since $\delta^2\tau$ involves both the independent and the dependent variables, the position of P depends upon the particular function τ which we are studying; e.g., upon the initially selected value $\tau(0)$. It is not immediately obvious that any such P exists, nor that there may not be more than one.

A single-valued function of a complex variable δ, with isolated singular points, can be represented [1] by power series as follows: In the δ-plane, draw concentric circles, one through each singular point, and all centered at the origin. Within the smallest circle, the function is represented by an ascending series in δ (i.e., there is a constant term, and a series of all positive integral powers of δ). This series is absolutely convergent and represents the given function for all values of δ within the circle. Outside the smallest circle, it diverges. On the circle itself, it may or may not converge. This circle is called the "circle of convergence" for that series; as a special case, its radius may be zero.

Outside the largest circle, the function is represented by a descending series, namely a series consisting of a constant and negative integral powers of δ. It is absolutely convergent outside the circle, divergent inside it, and may or may not converge on it.

In the region between any pair of circles, the function is represented by a series containing all integral powers of δ, both positive and negative, and a different series is required for each inter-circle region. Such a series is absolutely convergent in the region between the two circles to which it appertains, may or may not converge on those circles, and diverges outside the specified region.

All of these series together represent the given function at all points of the complex plane, except perhaps those on the various circles. The function exists for all points except the singular points themselves; for other points on the various circles, one of the series may or may not represent the function.

For the function in which we are interested, we at first use the hypothesis that there is one singular point, and that therefore there are two series: The ascending and descending ones, which have a common circle of convergence. A few terms of each series are given by Bremmer [2], and a greater number of terms are given by Johler, Walters, and Lilley [3].

Although one of the two series converges for any value of δ that does not lie on the common circle of convergence, the rate of convergence is slower the nearer the point is to that circle. For points very near the circle, the converging series will not, in general, provide a practical method of computation. Even within the limits of practical computation, it will be impossible to select any fixed number of terms at which to stop the computation, if given accuracy of the computed result is desired. Rather, the number of terms must be increased as δ gets nearer the circle.

One could, of course, always compute a fixed number of terms; but there would then be no assurance as to how accurate the result would be, nor as to which series should be used, nor even that either one converged. There is thus need for a formulation of the series which does not rely on expressions with a fixed number of terms.

2. Ascending Series

In 1955, the present author programed radio applications of (1) on the National Bureau of Standards Electronic Automatic Computer (SEAC). Some of the results have been published [4], but not the formulas. Instead of using a fixed number of terms of the series, recurrence formulas were developed for expressing each term as a quadratic function of the preceding ones. To derive the formulas for the ascending series, we express τ as a series in δ:

$$\tau=\sum_{n=0}^{\infty} b_n\delta^n. \tag{3}$$

Rewrite Riccati's equation as follows:

$$\frac{1}{d\tau/d\delta}-2\delta^2\tau+1=0. \tag{4}$$

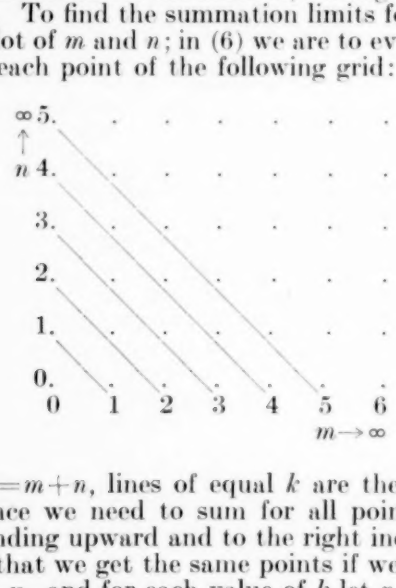
Insert into (4) the value of τ given in (3), and in one term change the summation index from n to m:

$$\frac{1}{\sum_{n=0}^{\infty} nb_n\delta^{n-1}}-2\sum_{m=0}^{\infty} b_m\delta^{m+2}+1=0. \tag{5}$$

Now clear of fractions, remembering that two absolutely convergent series may be multiplied term by term:

$$1-2\sum_{m=0}^{\infty}\sum_{n=0}^{\infty} nb_nb_m\delta^{m+n+1}+\sum_{n=0}^{\infty} nb_n\delta^{n-1}=0. \tag{6}$$

In the second term, introduce a new summation index k, and eliminate the index m, k being defined by $m=k-n$. To find the summation limits for n and k make a plot of m and n; in (6) we are to evaluate the term for each point of the following grid:



Since $k=m+n$, lines of equal k are the diagonal lines. Since we need to sum for all points of the grid, extending upward and to the right indefinitely, it is seen that we get the same points if we let k run from 0 to ∞, and for each value of k let n run from 0 to k.

$$1-2\sum_{k=0}^{\infty}\sum_{n=0}^{k} nb_nb_{k-n}\delta^{k+1}+\sum_{n=0}^{\infty} nb_n\delta^{n-1}=0. \tag{7}$$

Now change the summation indices again. In the second term, replace k by $j-2$, and in the last term replace n by j. Then write out explicitly the terms for $j=0$ and 1, and collect terms having a common summation:

$$1+b_1+\sum_{j=2}^{\infty}\left[jb_j-2\sum_{n=0}^{j-2} nb_nb_{j-n-2}\right]\delta^{j-1}=0. \tag{8}$$

Since this equation is to be satisfied for every value of δ within the circle of convergence, the coefficient of each power of δ must vanish. This imposes no condition on b_0, which is equal to $\tau(0)$, which we have previously proposed as the constant of integration. From the constant term of (8),

$$b_1=-1 \tag{9}$$

and for the higher powers of δ, we see that the bracket of (8) must be zero for every value of $j\geqq 2$, i.e.,

$$b_j=(2/j)\sum_{n=0}^{j-2} nb_nb_{j-n-2}. \tag{10}$$

In particular, $b_2=0$.

Now break the right side of (10) into two equal sums, each having a coefficient of $1/j$; and in one of them replace n by a new summation index m defined by $n=j-2-m$:

$$b_j=\frac{1}{j}\sum_{n=0}^{j-2} nb_nb_{j-n-2}+\frac{1}{j}\sum_{m=j-2}^{0}(j-2-m)b_{j-2-m}b_m. \tag{11}$$

Now replace summation-index m by n. Since the summations run over the same values of n, and since terms with the same n have also the same b-product, the series can be combined to give:

$$b_j=\frac{j-2}{j}\sum_{n=0}^{j-2} b_nb_{j-n-2} \text{ for } j\geqq 2. \tag{12}$$

Equations (9) and (12) show the coefficients to be used in (3), after the summation-index of the latter has been changed to j.

3. Recurrence Formulas for the Terms

For the SEAC computation, it is better to use recurrence formulas involving the terms rather than the coefficients. Let B_j represent a term of our series; i.e.,

$$B_j=b_j\delta^j. \tag{13}$$

Substituting into (3), (9), and (12), we get

$$\left.\begin{aligned}&\tau=\sum_{j=0}^{\infty} B_j;\ B_1=-\delta;\ B_2=0;\\ &B_j=\frac{j-2}{j}\delta^2\sum_{n=0}^{j-2} B_nB_{j-n-2} \qquad \text{for} \quad j\geqq 3,\end{aligned}\right\} \tag{14}$$

where $B_0=b_0=\tau(0)$ is the constant of integration, and may be chosen arbitrarily. In practice, it is a boundary condition, chosen to fit the conditions of the problem at hand.

4. Descending Series

Using a similar but somewhat more complicated method, we get for the descending series

$$\left.\begin{aligned} &\tau=\sum_{j=0}^{\infty} A_j;\ A_1=\left(-\frac{1}{2A_0}\right)\left(\frac{1}{\delta}\right);\\ &A_j=\left(-\frac{1}{2A_0}\right)\left[\sum_{n=1}^{j-1} A_n A_{j-n}\right.\\ &\qquad\left.-\frac{j-2}{j}\frac{1}{\delta^2}A_{j-2}\right] \quad \text{for} \quad j\geqq 2, \end{aligned}\right\} \tag{15}$$

where $A_0=\tau(\infty)$ is the constant of integration.

In order to determine which value of A_0 belongs to the same function $\tau(\delta)$ as does a given value of B_0, so that the results of the two series may be combined into a single function, other information about the problem at hand is used [3, pp. 1 to 4].

5. The SEAC Computations

The SEAC code used (14) and (15) as follows: One of them was selected more or less arbitrarily. The terms A_j or B_j were computed and saved, since each involves those which come before. When possible, computations were continued until a term was found which was negligibly small. (Real and imaginary parts of A_j or B_j each smaller than, say, 10^{-6}.) This determination began with $j=3$, since B_2 is identically zero. When such a term was reached, the computations were discontinued, and the A's or B's were summed.

If any A_j or B_j (starting with A_2 or B_3) was very large, that computation was abandoned, and the other formula was tried. The same thing was done if no negligibly small term was found within the space available for storing terms in the memory (about 65 terms could be stored). If, by the same criteria, the second formula did not converge, the computation for that particular value of $\tau(0)$ was abandoned. Some of the gaps in the published results [4] come from this cause. (Some occurred because of lack of time to complete the lengthy computations that had been planned.) This might occur either because the point was very near the circle of convergence, so that still more terms were needed; or because for certain functions there are more than one singular point, giving a region in which both series actually diverge.

6. Convergence of the Series

The criterion used for terminating the computations (real and imaginary parts of some term both less than 10^{-6}) is not rigorous, and under some circumstances it would be incorrect. This may be most easily studied by substituting into (12) so that each b_j is a polynomial in $b_0=B_0$, to give the equations of [3]. If B_0 happens to be such that any b_j is zero, then B_j will be zero regardless of δ, even though subsequent values of B might be of substantial size. For example, two out of every three b's are zero if $B_0=0$; or b_7 is zero if $5+8B_0^3=0$, etc.

For the problem under consideration, each value of B_0 for which $\tau(\delta)$ was to be derived had an amplitude of $\pi/3$, and the moduli were all transcendental numbers exceeding 1.8 [3, p. 20]. Being transcendental, they could not make any b_j exactly zero; and it is extremely unlikely that, among the small number of values that were used, there would be one that would make any b_j negligibly small. The expansions of b_j through $j=11$ as polynomials in b_0 [3] show that the coefficients of powers of b_0 tend to increase as j increases; and since the modulus of b_0 exceeds unity, it seems most probable that the values of b_j continue to increase as j increases. Hence, it seems warranted to conclude that when any B_j becomes negligibly small it is probably because $|\delta|$ is less than unity, and $|\delta^j|$ is small; so that succeeding values of $|B_j|$ probably will be still smaller, on account of the factor δ^2 in (14).

A similar argument seems to apply to the descending series. More reliable criteria would be to require that each of several consecutive terms of the series be negligibly small.

7. Bridging the Gap

Although neither series can be used for practical computation for points near the circle of convergence (or between the circles of convergence, if there is more than one singular point), the function itself is regular in any region not including a singular point; and for a point far from a singular point, the derivatives would not be excessive. Hence, if we are far from a singular point, it may be entirely proper to interpolate across the gap where neither series could be used.

Another method of getting values in this gap is by means of analytic continuation. We can use the ascending series to give a value of τ for some value of δ which we may call q. It is then possible [1, p. 196] to represent τ as an ascending series in $\delta-q$, which will be valid within the circle centered at q and passing through the nearest singular point. Such a series, obtained by methods analogous to those already described, is

$$\left.\begin{aligned} &\tau=\sum_{j=0}^{\infty} C_j;\ C_0=\tau(q);\ C_1=\frac{\delta-q}{2C_0q^2-1};\\ &C_j(1-2C_0q^2)=\frac{j-2}{j}(\delta-q)^2\sum_{n=0}^{j-2} C_nC_{j-n-2}\\ &\quad+2q(\delta-q)\frac{j-1}{j}\sum_{n=0}^{j-1} C_nC_{j-n-1}+q^2\sum_{n=1}^{j-1} C_nC_{j-n} \end{aligned}\right\} \tag{16}$$

No computations have actually been made with this formula.

8. References

[1] See any treatise on theory of a complex variable; e.g., Functions of a complex variable, by Édouard Goursat, translated by Earle Raymond Hedrick and Otto Dunkel, pp. 19, 80–84 (Ginn & Co., 1916).

[2] H. Bremmer, Terrestrial radio waves. Elsevier, New York, N.Y., 1949, p. 45. Bremmer errs in saying that there exists an intermediate region in which τ may be calculated from either series.

[3] J. R. Johler, L. C. Walters, and C. M. Lilley, Low- and very low-radiofrequency tables of ground wave parameters for the spherical earth theory: the roots of Riccati's differential equation (supplementary numerical data for NBS circular 573), NBS Tech. Note No. 7, pp. 5–6 (Feb. 1, 1959).

[4] James R. Wait and H. Herbert Howe, Amplitude and phase curves for ground-wave propagation in the band 200 cycles per second to 500 kilocycles, NBS Circ. 574 (1956).

Boulder, Colo.

(Paper 64B2–26)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 64B, No. 2, April–June 1960

On a Generalization of the Index Notation for Absolute Tensors of Arbitrary Order

Edmund H. Brown

(January 7, 1960)

A generalized vector index notation is introduced which facilitates study of properties of tensors regardless of order and variance of components. This generality is obtained by replacing the set of m indices, each of which vary in a single domain of n integers, by one *vector* index which varies over the integral lattice points of an m-dimensional domain. The notation also suggests an elegant treatment of order transformations of tensors.

1. Introduction

The usual index notation for the components of a tensor in a given coordinate system has the advantage that repetitious display of similar components of tensor quantities or equations is avoided without sacrificing directness and simplicity of algebraic manipulations. At the same time, retention of a single base symbol (to which the indices are attached) for each tensor is conceptually appealing since there is a one-to-one correspondence between base symbols and covariant objects. This correspondence is especially desirable in the expression of physical theories since physical observables are required to have the covariance property. A disadvantage of the notation is that, in the study of general properties of tensors, equations such as the transformation relations, the expressions for covariant derivatives, or the integrability conditions do not retain the same form, but require an increasing number of terms or factors with increasing tensor *order*. In this note, an amusing generalization of the usual notation is presented which permits a study of such general tensor properties in a form which is invariant with respect to order. In addition, this generalization suggests some new concepts, such as that of the *order transformation*.

2. Index Notation

Let T be an mth order tensor defined on an n-dimensional coordinate-space G and x^i be an arbitrary coordinate system in G. Then, m indices j_a may be defined in the domain $J(_1)$ consisting of the set of positive integers such that $0<j_a\leq n$ for $a=1,2,\ldots m$. By assigning one component of T in the coordinate system x^i to each set of the indices j_a the tensor T may be completely specified and written in the index notation as $t^{j_1 j_2 \cdots j_m}$.

For every coordinate transformation in G given by a set of n analytic functions $x^s=x^s(x^i)$ there will be a transformation of the components of the tensor T at a point x^i given by

$$t^{s_1 s_2 \cdots s_m}=t^{j_1 j_2 \cdots j_m} x^{s_1}_{j_1} x^{s_2}_{j_2} \cdots x^{s_m}_{j_m}, \tag{1.m}$$

where

$$x^{s_a}_{j_a}=\frac{\partial x^{s_a}}{\partial x^{j_a}}.$$

Thus, for a vector, $m=1$, the transformation law will be

$$t^s(x^r)=t^j(x^i)x^s_j; \tag{1.1}$$

and, in general, the transformation law of an mth order tensor will involve exactly m of the transformation matrix factors x^i_r.

If a tensor field $T(G)$ is defined on all points of G, the true rate of change of T is given by the covariant derivative

$$t^{j_1 j_2 \cdots j_m}|_k = \frac{\partial t^{j_1 j_2 \cdots j_m}}{\partial x^k} + t^{h j_2 \cdots j_m} \Gamma^{j_1}_{\cdot hk} + t^{j_1 h \cdots j_m} \Gamma^{j_2}_{\cdot hk} + \cdots + t^{j_1 j_2 \cdots h} \Gamma^{j_m}_{\cdot hk}, \qquad (2.\mathrm{m})$$

where the convenient vertical bar notation of Green and Zerna[1] is used. For a vector, the expression for the covariant derivative becomes merely

$$t^i|_k = \frac{\partial t^i}{\partial x^k} + t^h \Gamma^i_{\cdot hk}; \qquad (2.1)$$

and, in general, the covariant derivative of an mth order tensor will involve exactly m "corrective" terms in the Christoffel Connection $\Gamma^i_{\cdot jk}$.

As a final example, the integrability conditions for an mth order tensor may be written

$$t^{i_1 i_2 \cdots i_m}|_{kj} - t^{i_1 i_2 \cdots i_m}|_{jk} = t^{h i_2 \cdots i_m} R^{i_1}_{\cdot h|jk} + t^{i_1 h \cdots i_m} R^{i_2}_{\cdot h|jk} + \cdots + t^{i_1 i_2 \cdots h} R^{i_m}_{\cdot h|jk}, \qquad (3.\mathrm{m})$$

where $R^i_{j|kh}$ is the Riemann-Christoffel curvature tensor. For a vector, the integrability condition reduces to

$$t^i|_{kj} - t^i|_{jk} = t^h R^i_{\cdot h|jk}; \qquad (3.1)$$

and, in general, there will be exactly m terms on the right side of the integrability conditions (3.m) for every mth order tensor.

3. Index Vector

In the previous section, exactly m indices j_a were adjoined to the base symbol of an mth order tensor, with each of the j_a varying in a single domain $J_{(1)}$. We now consider the converse possibility of a Cartesian lattice index space $J_{(m)}$ which is the mth Cartesian product of m of the spaces $J_{(1)}$ and define a Cartesian *index vector* $\mathbf{j}$ in this space. For the moment, we consider only index vectors whose components are *positive* integers, that is, we restrict $J_{(m)}$ to the first hyperquadrant. For an mth order tensor on an n-dimensional space, a *maximum index vector* $\hat{\mathbf{j}}$ can be defined by the relations $\hat{j}_a = n$, $a = 1, \ldots, m$, for its components $\hat{j}_a$. Then, since there will be exactly n^m index vectors $\mathbf{j}$ which satisfy the conditions $0 < j_a \leq \hat{j}_a$, one component of the tensor may be assigned to each of index vectors $\mathbf{j}$ and the tensor itself represented by the symbol $t^{\mathbf{j}}$.

One of the simplest methods of forming higher-order tensors is by adjunction of vectors. Since coordinate *differentials* are vectors, a first application of this index vector notation results in the symbol $dx^{\mathbf{j}} = dx^{j_1} dx^{j_2} \ldots dx^{j_m}$. Although the transformation matrices x^r_j do not represent tensors, the symbol $x^{\mathbf{s}}_{\mathbf{j}} = x^{s_1}_{j_1} x^{s_2}_{j_2} \ldots x^{s_m}_{j_m}$ may be constructed by analogy with $dx^{\mathbf{j}}$. Using this generalized matrix, the transformation law for a tensor of *arbitrary order* may be written

$$t^{\mathbf{s}} = t^{\mathbf{j}} x^{\mathbf{s}}_{\mathbf{j}}, \qquad (1)$$

where (corresponding to the usual summation convention) the diagonal repetition of $\mathbf{j}$ indicates summation over all the index vectors in $J_{(m)}$.

By analogy with the derivation of Brillouin,[2] generalized expressions for the Christoffel Connection may be obtained by considering the displacement of a tensor of arbitrary order. If x^a are normal coordinates, that is, coordinates such that the covariant differential $Dt^{\mathbf{a}}$ is equal to the ordinary differential $dt^{\mathbf{a}}$ in the infinitesimal neighborhood of the point x^b, then

$$t^{\mathbf{a}}(x^b + dx^b) = t^{\mathbf{a}}(x^b) + Dt^{\mathbf{a}}.$$

[1] A. E. Green and W. Zerna, Theoretical elasticity (Oxford University Press, Oxford, England, 1954).
[2] L. Brillouin, Les tenseurs en mécanique et en élasticité (Dover Publications, New York, N.Y., 1946).

Transforming to arbitrary coordinates x^i by means of the transformation matrices $x^{\mathbf{i}}_{\mathbf{a}}(x^b+dx^b)$ and $x^{\mathbf{i}}_{\mathbf{a}}(x^b)$ gives

$$t^{\mathbf{i}}(x^k)=t^{\mathbf{a}}(x^b)x^{\mathbf{i}}_{\mathbf{a}}(x^b), \quad t^{\mathbf{i}}(x^k+dx^k)=\{t^{\mathbf{a}}(x^b)+Dt^{\mathbf{a}}\}x^{\mathbf{i}}_{\mathbf{a}}(x^b+dx^b)$$

$$=\{t^{\mathbf{a}}(x^b)+Dt^{\mathbf{a}}\}\{x^{\mathbf{i}}_{\mathbf{a}}(x^b)+x^{\mathbf{i}}_{\mathbf{a}b}dx^b\}=t^{\mathbf{i}}(x^k)+Dt^{\mathbf{i}}+t^{\mathbf{j}}x^{\mathbf{i}}_{\mathbf{a}b}x^{\mathbf{a}}_{\mathbf{j}}x^b_k dx^k,$$

where $x^{\mathbf{i}}_{\mathbf{a}b}=\partial x^{\mathbf{i}}_{\mathbf{a}}/\partial x^b$ and where we neglect the term containing the second-order differential, $Dt^{\mathbf{a}}x^{\mathbf{i}}_{\mathbf{a}b}dx^b$. Defining the Christoffel Connection by $\Gamma^{\mathbf{i}}_{\cdot\mathbf{j}k}=-x^{\mathbf{i}}_{\mathbf{a}b}x^{\mathbf{a}}_{\mathbf{j}}x^b_k$, the covariant differential can be written

$$Dt^{\mathbf{i}}=dt^{\mathbf{i}}+t^{\mathbf{j}}\Gamma^{\mathbf{i}}_{\cdot\mathbf{j}k}dx^k,$$

and the generalization of the expression (2.m) for the covariant derivative of a tensor of *arbitrary order* becomes

$$t^{\mathbf{i}}|_k=\frac{\partial t^{\mathbf{i}}}{\partial x^k}+t^{\mathbf{j}}\Gamma^{\mathbf{i}}_{\cdot\mathbf{j}k}. \tag{2}$$

If the normal coordinates x^a are assumed to be obtained by transformation from a third set of arbitrary coordinates x^r, we can write

$$\Gamma^{\mathbf{i}}_{\cdot\mathbf{j}k}=-(x^{\mathbf{i}}_{\mathbf{r}}x^{\mathbf{r}}_{\mathbf{a}})_b x^{\mathbf{a}}_{\mathbf{j}}x^b_k=-x^{\mathbf{r}}_{\mathbf{a}b}x^{\mathbf{a}}_{\mathbf{p}}x^b_s x^{\mathbf{i}}_{\mathbf{r}}x^{\mathbf{p}}_{\mathbf{j}}x^s_k-x^{\mathbf{i}}_{\mathbf{r}s}x^{\mathbf{r}}_{\mathbf{a}}x^{\mathbf{a}}_{\mathbf{j}}x^s_b x^b_k$$

or,

$$\Gamma^{\mathbf{i}}_{\cdot\mathbf{j}k}=\Gamma^{\mathbf{r}}_{\cdot\mathbf{p}s}x^{\mathbf{i}}_{\mathbf{r}}x^{\mathbf{p}}_{\mathbf{j}}x^s_k-x^{\mathbf{i}}_{\mathbf{r}s}x^{\mathbf{r}}_{\mathbf{j}}x^s_k,$$

which is the generalization of the well-known Christoffel transformation law. Clearly, the original definition of $x^{\mathbf{a}}_{\mathbf{j}}$ requires that

$$t^{\mathbf{j}}\Gamma^{\mathbf{i}}_{\cdot\mathbf{j}k}=t^{j i_2\cdots i_m}\Gamma^{i_1}_{\cdot jk}+t^{i_1 j\cdots i_m}\Gamma^{i_2}_{\cdot jk}+\cdots+t^{i_1 i_2\cdots j}\Gamma^{i_m}_{\cdot jk},$$

by the law of differentiation of products and the relation $\delta^i_j=x^i_a x^a_j$.

For an arbitrary tensor field $t^{\mathbf{i}}(x^k)$, the expression for the covariant differential permits a calculation of the change in the components at a fixed point after integration around a given curve C of G. If $\delta t^{\mathbf{i}}$ is the change at the point, then

$$\delta t^{\mathbf{i}}=\oint_C dt^{\mathbf{i}}=\oint_C Dt^{\mathbf{i}}-\oint_C t^{\mathbf{h}}\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}k}dx^k=\oint_C t^{\mathbf{i}}|_k dx^k-\oint_C t^{\mathbf{h}}\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}k}dx^k$$

$$=\frac{1}{2}\iint_S\left\{\frac{\partial t^{\mathbf{i}}|_k}{\partial x^j}-\frac{\partial t^{\mathbf{i}}|_j}{\partial x^k}\right\}ds^{jk}-\frac{1}{2}\iint_S\left\{\frac{\partial t^{\mathbf{h}}\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}k}}{\partial x^j}-\frac{\partial t^{\mathbf{h}}\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}j}}{\partial x^k}\right\}ds^{jk}$$

$$=-\frac{1}{2}\iint_S\{t^{\mathbf{i}}|_{jk}-t^{\mathbf{i}}|_{kj}\}ds^{jk}-\frac{1}{2}\iint_S t^{\mathbf{h}}\left\{\left(\frac{\partial\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}k}}{\partial x^j}+\Gamma^{\mathbf{i}}_{\cdot\mathbf{l}j}\Gamma^{\mathbf{l}}_{\cdot\mathbf{h}k}\right)-\left(\frac{\partial\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}j}}{\partial x^k}+\Gamma^{\mathbf{i}}_{\cdot\mathbf{l}k}\Gamma^{\mathbf{l}}_{\cdot\mathbf{h}j}\right)\right\}ds^{jk},$$

where S is the surface enclosed by C and ds^{jk} the differential surface element. Following the usual definition of the curvature tensor,

$$R^{\mathbf{i}}_{\cdot\mathbf{h}|jk}=\left(\frac{\partial\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}k}}{\partial x^j}+\Gamma^{\mathbf{i}}_{\cdot\mathbf{l}j}\Gamma^{\mathbf{l}}_{\cdot\mathbf{h}k}\right)-\left(\frac{\partial\Gamma^{\mathbf{i}}_{\cdot\mathbf{h}j}}{\partial x^k}+\Gamma^{\mathbf{i}}_{\cdot\mathbf{l}k}\Gamma^{\mathbf{l}}_{\cdot\mathbf{h}j}\right),$$

so that the change $\delta t^{\mathbf{i}}$ may be written

$$\delta t^{\mathbf{i}}=-\frac{1}{2}\iint_S\{t^{\mathbf{i}}|_{jk}-t^{\mathbf{i}}|_{kj}+t^{\mathbf{h}}R^{\mathbf{i}}_{\cdot\mathbf{h}|jk}\}ds^{jk}.$$

Since integrability requires $\delta t^{\mathbf{i}}=0$ for arbitrary area S, we arrive at the generalization of the integrability conditions (3.m) in the form

$$t^{\mathbf{i}}{}_{|kj}-t^{\mathbf{i}}{}_{|jk}=t^{\mathbf{h}}R^{\mathbf{i}}_{\cdot\mathbf{h}|jk}. \tag{3}$$

Thus, to be integrable, a tensor field cannot be completely arbitrary, but must satisfy the conditions (3). In the case of a tensor field obtained by *parallel* displacement to all points of a tensor given initially at one point, we must have $Dt^{\mathbf{i}}\equiv 0$; thus, the field can be integrable only if the space is Euclidean, $\mathrm{R}^{\mathbf{i}}_{\cdot\mathbf{h}|jk}\equiv 0$.

4. Mixed Index Vectors

The index vector notation facilitates study of general tensor properties regardless of order. In such a study, however, specification of the covariancy or contravariancy of components with respect to particular indices may be even less important, especially in a metric space where components may be immediately transformed from one type of variance to the other. Thus, a general notation should allow interpretation of such differences in variance without actually displaying the differences in its symbols.

Let us consider what significance can be attached to an index space $J_{(m)}$ which includes all hyperquadrants, that is, which allows both positive and negative integers as components of the index vector $\mathbf{j}$. Each hyperquadrant in the enlarged $J_{(m)}$ will be bounded by m hyperplanes formed on m *semi*-axes and, thus, there will be 2^m such hyperquadrants in $J_{(m)}$. Defining a maximum index vector $\hat{\mathbf{j}}$ by $\hat{j}_a=\pm n$ with some particular signature $\sigma=\{+,+,-,+\ \ldots\}$ for the signs of the components then, with the restrictions sign $j_a=$sign $\hat{j}_a$ and $0<|j_a|\leq|\hat{j}_a|$, there will be exactly n^m index vectors $\mathbf{j}$ for each maximum index vector $\hat{\mathbf{j}}$, that is, exactly the same number as components of an mth order tensor defined on an n-dimensional space. In addition, the number of maximum index vectors will be, clearly, equal to the total number of combinations of covariant and contravariant indices for an mth order tensor.

The above equality, thus, permits the following correspondence: To each contravariant index j_a we assign a *positive* sign to the respective component of $\hat{\mathbf{j}}$; to each covariant index j_b we assign a *negative* sign to the respective component of $\hat{\mathbf{j}}$; and, as the index vector $\mathbf{j}$ varies in $J_{(m)}$ subject to the conditions sign $j_a=$sign $\hat{j}_a$ and $0<|j_a|\leq|\hat{j}_a|$, we assign to each $\mathbf{j}$ one of the respective components of the mixed tensor corresponding to $\hat{\mathbf{j}}$. All the $2^m n^m$ index vectors in $J_{(m)}$ thus specify all the possible covariant, contravariant, and mixed components of a given mth order tensor defined on an n-dimensional coordinate space.

In particular, if $t^{\mathbf{i}}$ are components of arbitrary variance (that is, some of the components of the index vector may be negative) of a given tensor T, then $t^{-\mathbf{i}}=t_{\mathbf{i}}$ and the form of the transformation law (1) remains valid for this generalization if we also define $x^{-\mathbf{r}}_{-\mathbf{i}}=x^{\mathbf{i}}_{\mathbf{r}}$. Thus, (1) may be written as either $t^{\mathbf{r}}=t^{\mathbf{i}}x^{\mathbf{r}}_{\mathbf{i}}$ or $t_{\mathbf{s}}=t_{\mathbf{j}}x^{\mathbf{j}}_{\mathbf{s}}$ regardless of the variance of the components.

Defining $g_{\mathbf{ij}}=g_{i_1j_1}g_{i_2j_2}\ldots g_{i_mj_m}$, $g^{\mathbf{ij}}=g^{i_1j_1}g^{i_2j_2}\ldots g^{i_mj_m}$ and $\delta^{\mathbf{i}}_{\mathbf{j}}=g^{\mathbf{ik}}g_{\mathbf{kj}}=\delta^{i_1}_{j_1}\delta^{i_2}_{j_2}\ldots\delta^{i_m}_{j_m}$, where the indices i_a and j_a may be either positive or negative, the corresponding "covariant" Christoffel Connection will be given by $\Gamma_{\mathbf{i},jk}=g_{\mathbf{ih}}\Gamma^{\mathbf{h}}_{\mathbf{j}k}$. Using the identity

$$\frac{\partial g_{\mathbf{ik}}}{\partial x^j}=(\Gamma_{i_1k_1}+\Gamma_{k_1i_1})g_{i_2k_2}\ldots g_{i_mk_m}+\ldots+(\Gamma_{i_mjk_m}+\Gamma_{k_mji_m})g_{i_1k_1}\ldots g_{i_{m-1}k_{m-1}},$$

the expression (2) for the covariant derivative of a tensor of arbitrary order may be written in either of the forms

$$t^{\mathbf{r}}|_s=\frac{\partial t^{\mathbf{r}}}{\partial x^s}+t^{\mathbf{p}}\,\Gamma^{\mathbf{r}}_{\cdot\mathbf{p}s} \qquad \text{or} \qquad t_{\mathbf{i}}|_s=\frac{\partial t_{\mathbf{i}}}{\partial x^s}-t_{\mathbf{k}}\Gamma^{\mathbf{k}}_{\cdot\mathbf{i}s},$$

regardless of variance. The similar generalization of the integrability conditions (3) for tensors of arbitrary mixed variance is obvious.

5. Order Transformations

The identity $g_{ij}t^{j}=t^{-i}$ which follows from the results of the last section suggests the definition $(t^{i})^{-1}=t^{-i}$ or, generalizing the latter, the definition $t^{i}=(t^{j})^{\omega}=t^{\omega j}$, where $\boldsymbol{\omega}$ is a rotation operator in $\boldsymbol{J}_{(m)}$ subject to the restriction $0<|i_a|\leq n$. Thus, a variance transformation of the components of a tensor T on G can be considered equivalent to a restricted orthogonal transformation of $\hat{\mathbf{j}}$ in the index space $\boldsymbol{J}_{(m)}$.

By removing the restriction, a more general orthogonal transformation $\boldsymbol{\alpha}$ can be defined which can be called a decreasing *order transformation*, that is, $\boldsymbol{\alpha}$ will be a transformation which rotates $\hat{\mathbf{j}}$ into a subspace $\boldsymbol{J}_{(\nu)}$ of $\boldsymbol{J}_{(m)}$ and, correspondingly, transforms the tensor $T_{(m)}$ into $T_{(\nu)}$. If $\hat{\boldsymbol{\alpha}}$ is a *maximum* order transformation, that is, one which transforms $\hat{\mathbf{j}}$ into a zero vector at the origin of $\boldsymbol{J}_{(m)}$, then clearly there will be the same total number of such possible maximum order transformations $\boldsymbol{\alpha}_{(i)}$ as there are invariants of the tensor T.

In particular, the scalar $t^{0}=t^{i}t^{-i}=t^{i}t_{i}=t^{i_1 i_2\cdots i_n}t_{i_1 i_2\cdots i_n}$ is the only invariant possessed by tensors of *all* orders (and thus might be called the "fundamental quadratic invariant"). Just as *contraction* of a tensor is indicated in the ordinary notation by summation over a pair of indices (one covariant, one contravariant), contraction in this new notation results in a tensor $t^{k(\nu)}=\Sigma_{i_k,\,i_h}t^{i\,\nu)}$, subject to the restriction $i_k+i_h=0$, with k, h fixed, where $\nu=n-2$. The reduction in dimension of the index space results from, first, restricting the index vector to the hypersurface defined by $i_k+i_h=0$ and, second, summing the components of t^{i} (and assigning each sum to the corresponding sum of the index vectors) over the *values* of the index i_k; thus, the dimension of the index space is reduced by *two*.

Further consideration of general order transformations should be of considerable value in the theory of groups as used in quantum mechanics, in the study of irreducible tensors, and in determining the invariants of tensors of arbitrary order. If such general order transformations were unrestricted they would clearly lead to tensors of more general type, that is, tensors represented by a set of components defined, not as functions of a point in a single space, but as functions of M points in M spaces having possibly different dimensions.

BOULDER, COLO. (Paper 64B2–27)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 64B, No. 2, April–June 1960

Upper and Lower Bounds for the Center of Flexure

Lawrence E. Payne

(February 1, 1960)

There is disagreement in the literature as to how the center of flexure of an isotropic elastic beam should be defined. In either of the two most widely accepted definitions, upper and lower bounds for the coordinates of the center of flexure are obtained.

1. Introduction

We consider a cantilever beam of uniform cross section and choose the z-axis to coincide with the line of cross-section centroids. The plane end $z=0$ is assumed fixed and the other end $z=l$ is loaded by some distribution of forces which is statically equivalent to a single force with components $(W_x, W_y, 0)$ acting at a load point (x_L, y_L, l). In the treatment of the flexure problem [9] [1] (a similar treatment appears in [3]) we seek a load point (x_f, y_f, l) which has the property that any load $(W_x, W_y, 0)$ applied through this point produces no local twist at the centroid of the section, i.e., the mean value of the local twist over the cross section vanishes (see [9] p. 200). This point (x_f, y_f, l) is referred to as the center of flexure [9]. The determination of (x_f, y_f, l) permits us to break up the general flexure problem into two separate problems—one of pure torsion and one of pure flexure. This point is sometimes referred to as the center of shear (see, for instance, [3]), and it is well known that there is difference of opinion in the literature as to the way in which the point should be defined (see, for instance, [1, 2, 4, 10, 11]. A discussion of these differences has been given in a recent paper by Pearson [7], who demonstrates that in general the two expressions for the center of shear yield results which are very near to one another.

It is not our point here to enter into the discussion of the relative merits of the two definitions. We merely point out that the expression of Trefftz [10] is simpler and independent of Poisson's ratio. The center of flexure of [9], on the other hand, permits an easy decomposition of the general flexure problem.

In order to avoid confusion we shall hereafter refer to the point defined in [9] as the center of flexure and that given by Trefftz [10] as the center of shear.

In this paper then we derive upper and lower bounds for the coordinates of the center of flexure (x_f, y_f, l) and the center of shear (x_s, y_s, l). In the first case we use the formulation of the flexure problem derived in [9] and thus concern ourselves with the problem of pure flexure.

It is of course well known (see, for instance, [7]) that if we have the solution to the pure torsion problem for the beam in question then we can determine both the center of shear and the center of flexure without knowledge of the solution to the flexure problem. As would be expected it is possible in each case to obtain upper and lower bounds for the coordinates of the point by approximating the solution to the torsion problem alone.

2. Bounds for the Center of Flexure

We denote by $\mathscr{D}$ the cross section of the beam and let C be its boundary. (The region may be simply or multiply connected.) The origin of a rectangular coordinate system is taken at the centroid of $\mathscr{D}$. For simplicity the x and y axes are chosen to be principal axes of inertia, i.e.,

$$\iint_{\mathscr{D}} xy dA=0. \tag{2.1}$$

[1] Figures in brackets indicate the literature references at the end of this paper.

According to [9] the pure flexure problem is solved once the solutions to the following boundary value problems have been determined:

$$\frac{\partial^2\varphi_1}{\partial x^2}+\frac{\partial^2\varphi_1}{\partial y^2}=0, \quad \frac{\partial^2\varphi_2}{\partial x^2}+\frac{\partial^2\varphi_2}{\partial y^2}=0 \qquad \text{in } \mathscr{D} \tag{2.2}$$

$$\frac{\partial\varphi_1}{\partial n}=[(1+\sigma)x^2-\sigma y^2]n_x \tag{2.3}$$

$$\frac{\partial\varphi_2}{\partial n}=[(1+\sigma)y^2-\sigma x^2]n_y \qquad \text{on } C$$

where σ is Poisson's ratio and n_x and n_y are the x and y components of the unit normal on C. (The normal is assumed directed outward from $\mathscr{D}$.) The coordinates of the center of flexure (x_f, y_f) are defined by:

$$x_f=\frac{1}{2(1+\sigma)I_x}\iint_{\mathscr{D}}[x\frac{\partial\varphi_2}{\partial y}-y\frac{\partial\varphi_2}{\partial x}-(1+\sigma)xy^2+\sigma x^3]dA, \tag{2.4}$$

$$y_f=\frac{1}{2(1+\sigma)I_y}\iint_{\mathscr{D}}[x\frac{\partial\varphi_1}{\partial y}-y\frac{\partial\varphi_1}{\partial x}+(1+\sigma)x^2y-\sigma y^3]dA \tag{2.5}$$

where

$$I_x=\iint_{\mathscr{D}}y^2dA \tag{2.6}$$

$$I_y=\iint_{\mathscr{D}}x^2dA. \tag{2.7}$$

Thus

$$2(1+\sigma)I_xx_f=\iint_{\mathscr{D}}\left[\frac{\partial}{\partial y}(x\varphi_2)-\frac{\partial}{\partial x}(y\varphi_2)-(1+\sigma)xy^2+\sigma x^3\right]dA. \tag{2.8}$$

Using the divergence theorem we have

$$2(1+\sigma)I_xx_f=\oint_C\varphi_2(xn_y-yn_x)ds-J \tag{2.9}$$

where

$$J=\iint_{\mathscr{D}}[(1+\sigma)xy^2-\sigma x^3]dA. \tag{2.10}$$

Since J is a geometric quantity it can be computed explicitly.

We note now that the coefficient of φ_2 in the boundary integral in (2.9) is proportional to the boundary value of the normal derivative of the warping function φ in the pure torsion problem. In fact

$$\frac{\partial\varphi}{\partial n}=yn_x-xn_y \qquad \text{on } C. \tag{2.11}$$

Thus if we let

$$\chi=2(1+\sigma)I_xx_f+J \tag{2.12}$$

we find

$$\chi=-\oint_C\varphi_2\frac{\partial\varphi}{\partial n}ds=-D(\varphi,\varphi_2) \tag{2.13}$$

where

$$D(\varphi,\varphi_2)=\iint_{\mathscr{D}}\left[\frac{\partial\varphi}{\partial x}\frac{\partial\varphi_2}{\partial x}+\frac{\partial\varphi}{\partial y}\frac{\partial\varphi_2}{\partial y}\right]dA. \tag{2.14}$$

The determination of bounds for x_f is thus equivalent to the determination of bounds for χ.

Note that if the warping function φ is known the value of x_f can be determined explicitly since

$$\chi=-\oint_C \varphi[(1+\sigma)y^2-\sigma x^2]n_y ds. \tag{2.15}$$

If φ is not known we introduce two arbitrary harmonic functions u and u_2 and define

$$a_1=-D(\varphi,u_2)=-\oint_C u_2(yn_x-xn_y)ds, \tag{2.16}$$

$$a_2=-D(u,\varphi_2)=-\oint_C u[(1+\sigma)y^2-\sigma x^2]n_y ds, \tag{2.17}$$

$$a_3=-D(u,u_2). \tag{2.18}$$

With u and u_2 prescribed we can determine a_1, a_2, and a_3 explicitly. We now form the expression

$$\begin{aligned}\chi-a_1-a_2+a_3&=-D(\varphi-u,\varphi_2-u_2)\\&=-\oint_C(\varphi-u)\left\{[(1+\sigma)y^2-\sigma x^2]n_y-\frac{\partial u_2}{\partial n}\right\}ds.\end{aligned} \tag{2.19}$$

An application of Schwarz's inequality yields

$$(\chi-a_1-a_2+a_3)^2\leq\oint_C(\varphi-u)^2ds\oint_C\left\{[(1+\sigma)y^2-\sigma x^2]n_y-\frac{\partial u_2}{\partial n}\right\}^2ds. \tag{2.20}$$

It remains now to approximate the first boundary integral on the right of (2.20) in terms of known quantities.

By adding an appropriate constant (if necessary) it is possible to choose $\varphi-u$ in such a a way that

$$\oint_C(\varphi-u)ds=0. \tag{2.21}$$

With $\varphi-u$ so chosen we shall estimate the first integral on the right of (2.20) in terms of the boundary integral of the square of its normal derivative. To this end we note that

$$\frac{\oint_C\left(\frac{\partial}{\partial s}(\varphi-u)\right)^2ds}{\oint_C(\varphi-u)^2ds}\geq\min_{\oint_C vds=0}\frac{\oint_C\left(\frac{\partial v}{\partial s}\right)^2ds}{\oint_C v^2ds}=\frac{4\pi^2}{L^2} \tag{2.22}$$

where L denotes the length of the perimeter of C, $\partial/\partial s$ denotes the tangential derivative on C, and v is any function continuous on C and satisfying the condition $\oint_C vds=0$. The quantity on the right of (2.22) is just the eigenvalue of the vibrating string equation, the differential equation which arises as the Euler equation for the minimum of the quotient $\left(\oint_C(\partial v/\partial s)^2ds\right)/\oint_C v^2ds$. It follows then that

$$\oint_C(\varphi-u)^2ds\leq\frac{L^2}{4\pi^2}\oint_C\left[\frac{\partial}{\partial s}(\varphi-u)\right]^2ds. \tag{2.23}$$

We seek now an inequality which relates the integral on the right-hand side of (2.23) to the boundary integral of the square of the normal derivative of $\varphi-u$. We introduce therefore,

an arbitrary continuous vector function with components f_k and consider the generalized Rellich identity (see [6, 8]). Letting $w=\varphi-u$, we have

$$0=\iint_{\mathscr{D}} f_k \frac{\partial w}{\partial x_k}\,\Delta w\,dA=\oint_C f_k \frac{\partial w}{\partial x_k}\frac{\partial w}{\partial n}\,ds-\frac{1}{2}\oint_C f_k n_k|\text{grad } w|^2 ds-\iint_{\mathscr{D}}\frac{\partial f_k}{\partial x_l}\frac{\partial w}{\partial x_k}\frac{\partial w}{\partial x_l}\,dA+\frac{1}{2}\iint_{\mathscr{D}}\frac{\partial f_k}{\partial x_k}|\text{grad } w|^2 dA \tag{2.24}$$

where Δ denotes the Laplace operator ($\Delta=(\partial^2/\partial x^2)+(\partial^2/\partial y^2)$) and summation is to be carried out over the repeated indices $k,l=1,2$. The f_k are now assumed to be so chosen that

$$p=f_k n_k>0 \text{ on } C. \tag{2.25}$$

A decomposition of the derivatives of u in the boundary integrals into normal and tangential components yields (see [8])

$$\oint_C p\left(\frac{\partial w}{\partial s}\right)^2 ds=\oint_C p\left(\frac{\partial w}{\partial n}\right)^2 ds+2\oint_C f_k\frac{\partial x_k}{\partial s}\frac{\partial w}{\partial s}\frac{\partial w}{\partial n}\,ds+\iint_{\mathscr{D}}\frac{\partial f_k}{\partial x_k}|\text{grad } w|^2 dA-2\iint_{\mathscr{D}}\frac{\partial f_k}{\partial x_l}\frac{\partial w}{\partial x_k}\frac{\partial w}{\partial x_l}\,dA. \tag{2.26}$$

Since the functions f_k are prescribed, we can find a constant γ such that (see [6, p. 555])

$$\iint_{\mathscr{D}}\frac{\partial f_k}{\partial x_k}|\text{grad } w|^2 dA-2\iint_{\mathscr{D}}\frac{\partial f_k}{\partial x_l}\frac{\partial w}{\partial x_k}\frac{\partial w}{\partial x_l}\,da\le\gamma D(w,w). \tag{2.27}$$

We make use also of the ordinary Green's identity

$$D(w,w)=\oint_C w\frac{\partial w}{\partial n}\,ds, \tag{2.28}$$

which yields by Schwarz's inequality

$$D(w,w)\le\left(\oint_C w^2 ds\right)^{\frac{1}{2}}\left(\oint_C\left(\frac{\partial w}{\partial n}\right)^2 ds\right)^{\frac{1}{2}}. \tag{2.29}$$

In view of (2.23), eq (2.29) yields

$$D(w,w)\le\frac{L}{2\pi}\left[\oint_C\left(\frac{\partial w}{\partial s}\right)^2 ds\right]^{\frac{1}{2}}\left[\oint_C\left(\frac{\partial w}{\partial n}\right)^2 ds\right]^{\frac{1}{2}}. \tag{2.30}$$

We now insert (2.27) and (2.30) back into (2.26), make use of the Schwarz inequality and obtain

$$\oint_C p\left(\frac{\partial w}{\partial s}\right)^2 ds\le\oint_C p\left(\frac{\partial w}{\partial n}\right)^2 ds+2\left[\oint_C\frac{\left(f^k\frac{\partial x_k}{\partial s}\right)^2}{p}\left(\frac{\partial w}{\partial n}\right)^2 ds\right]^{\frac{1}{2}}\left[\oint_C p\left(\frac{\partial w}{\partial s}\right)^2 ds\right]^{\frac{1}{2}}$$
$$+\frac{\gamma L}{2\pi p_{\min}^{\frac{1}{2}}}\left[\oint_C p\left(\frac{\partial w}{\partial s}\right)^2 ds\right]^{\frac{1}{2}}\left[\oint_C\left(\frac{\partial w}{\partial n}\right)^2 ds\right]^{\frac{1}{2}}, \tag{2.31}$$

where $p_{\min}$ is the minimum value of p on C. This inequality yields

$$\left[\oint_C p\left(\frac{\partial w}{\partial s}\right)^2 ds\right]^{\frac{1}{2}}\le\left[\oint_C\frac{\left(f^k\frac{\partial x_k}{\partial s}\right)^2}{p}\left(\frac{\partial w}{\partial n}\right)^2 ds\right]^{\frac{1}{2}}+\left[\frac{\gamma^2L^2}{4\pi^2 p_{\min}}\oint_C\left(\frac{\partial w}{\partial n}\right)^2 ds\right]^{\frac{1}{2}}+\left\{\left(\left[\oint_C\frac{\left(f^k\frac{\partial x_k}{\partial s}\right)^2}{p}\left(\frac{\partial w}{\partial n}\right)^2 ds\right]^{\frac{1}{2}}\right.\right.$$
$$\left.\left.+\left[\frac{\gamma^2L^2}{4\pi^2 p_{\min}}\oint_C\left(\frac{\partial w}{\partial n}\right)^2 ds\right]^{\frac{1}{2}}\right)^2+\oint_C p\left(\frac{\partial w}{\partial n}\right)^2 ds\right\}^{\frac{1}{2}}. \tag{2.32}$$

From (2.32) it is thus always possible to obtain the inequality

$$\oint_C \left(\frac{\partial w}{\partial s}\right)^2 ds \leq \alpha \oint_C \left(\frac{\partial w}{\partial n}\right)^2 ds \tag{2.33}$$

where

$$\alpha p_{\min} = (\beta + [\beta^2 + p_{\max}]^{\frac{1}{2}})^2 \tag{2.34}$$

and

$$\beta = \left\{ \frac{\left| f^k \frac{\partial x_k}{\partial s} \right|}{p^{\frac{1}{2}}} \right\}_{\max} + \frac{\gamma L}{2\pi p_{\min}^{\frac{1}{2}}}. \tag{2.35}$$

If the boundary C is star-shaped with respect to some point (x_o, y_o) in $\mathscr{D}$, the problem is simplified considerably. The fact that the boundary is star-shaped with respect to (x_o, y_o) is equivalent to the condition that

$$q = (x - x_0) n_x + (y - y_0) n_y > 0 \tag{2.36}$$

at every point on C. Thus, in this case, (2.25) is satisfied if we let

$$f_1 = x - x_0, \qquad f_2 = y - y_0. \tag{2.37}$$

The left-hand side of (2.27) then vanishes identically, and hence we take $\gamma \equiv 0$ in (2.31), (2.32), and (2.35). We find in this case instead of (2.33) (see [5]) that

$$\oint_C \left(\frac{\partial w}{\partial s}\right)^2 ds \leq q_{\min}^{-1} \left[1 + \left|\frac{\partial r}{\partial s}\right|_{\max}\right]^2 \oint_C \frac{r^2}{q} \left(\frac{\partial w}{\partial n}\right)^2 ds. \tag{2.38}$$

In the general case then, an inequality of the form

$$\oint_C \left(\frac{\partial(\varphi - u)}{\partial s}\right)^2 ds \leq \oint_C A \left[\frac{\partial}{\partial n}(\varphi - u)\right]^2 ds \tag{2.39}$$

is obtained, where A is a completely determined positive function. Inserting (2.39) and (2.23) into (2.20), we find

$$(\chi - a_1 - a_2 + a_3)^2 \leq \frac{L_4^2}{4\pi^2} \oint_C A \left[\frac{\partial u}{\partial n} - y n_x + x n_y\right]^2 ds \oint_C \left\{\frac{\partial u_2}{\partial n} - [(1+\sigma)y^2 - \sigma x^2] n_y\right\}^2 ds. \tag{2.40}$$

(It is apparent that the quantity A may be placed in the second integral rather than in the first if it is desirable.) We now use the Rayleigh-Ritz technique to make the two integrals on the right of (2.40) small.

Note that it is possible to derive close upper and lower bounds for x_f without approximating the flexure function at all. In fact, if we use instead of u a conjugate function v defined by[2]

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \qquad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}, \tag{2.41}$$

we obtain (setting $u_2 \equiv 0$)

$$|\chi - a_2|^2 \leq \frac{L^2}{4\pi^2} \oint_C \left[\frac{\partial}{\partial s}(v - \tfrac{1}{2} r^2)\right]^2 ds \oint_C A \{[(1-\sigma)y^2 - \sigma x^2] n_y\}^2 ds. \tag{2.42}$$

In this case the A term has been put in the second integral since this expression is a purely geometrical quantity (independent of v) and can be computed explicitly.

[2] The function v is not to be confused with that used in (2.22).

In an entirely analogous way we find that

$$-2(1+\sigma)I_y y_f+\iint_{\mathscr{D}}[(1+\sigma)x^2y-\sigma y^3]dA=D(\varphi,\varphi_1). \tag{2.43}$$

By introducing the two arbitrary harmonic functions u and u_1 and proceeding as before, we obtain upper and lower bounds for y_f. Note that if we choose $u_1,u_2\equiv 0$ and merely approximate the warping function (or the conjugate function) on the right, then the same minimizing function u (or v) minimizes the error term in each case.

3. Bounds for the Center of Shear

We look now at the expression for the center of shear (x_s,y_s,l) as defined by Trefftz [10]. The coordinates x_s and y_s are given by

$$x_s=-\frac{1}{I_x}\iint_{\mathscr{D}}\varphi\, y dA \tag{3.1}$$

$$y_s=\frac{1}{I_y}\iint_{\mathscr{D}}\varphi\, x dA. \tag{3.2}$$

Equation (3.2) may be rewritten as

$$y_s=\frac{1}{6I_y}\iint_{\mathscr{D}}(\Delta x^3)\varphi\, dA. \tag{3.3}$$

By Green's formula we have

$$y_s=\frac{1}{6I_y}\oint_C\left[3x^2n_x\varphi-x^3\frac{\partial\varphi}{\partial n}\right]ds$$

$$=\frac{1}{6I_y}\oint_C[3x^2n_x\varphi-x^3(yn_x-xn_y)]\,ds. \tag{3.4}$$

The divergence theorem then yields

$$y_s+\frac{1}{2I_y}\iint_{\mathscr{D}}x^2y dA=\frac{1}{2I_y}\oint_C x^2n_x\varphi ds. \tag{3.5}$$

We introduce the notation

$$\Phi=y_s+\frac{1}{2I_y}\iint_{\mathscr{D}}x^2y dA, \tag{3.6}$$

and two harmonic functions h_1 and h_2. Let

$$b_1=\oint_C h_1x^2n_x ds, \tag{3.7}$$

$$b_2=\oint_C\varphi\frac{\partial h_2}{\partial n}ds=\oint_C h_2[yn_x-xn_y]ds, \tag{3.8}$$

$$b_3=\oint_C h_1\frac{\partial h_2}{\partial n}ds. \tag{3.9}$$

Combining (3.5) to (3.9) we obtain

$$\Phi - b_1 - b_2 + b_3 = \oint_C (\varphi - h_1)(x^2 n_x - h_2) ds. \tag{3.10}$$

Then by Schwarz's inequality

$$[\Phi - b_1 - b_2 - b_3]^2 \leq \oint_C (\varphi - h_1)^2 ds \oint_C \left(x^2 n_x - \frac{\partial h_2}{\partial n} \right)^2 ds. \tag{3.11}$$

Again it is desirable to choose $(\varphi - h_1)$ in such a way that

$$\oint_C (\varphi - h_1) ds = 0. \tag{3.12}$$

Using (2.23) and (2.39), we obtain finally

$$[\Phi - b_1 - b_2 + b_3]^2 \leq \frac{L^2}{4\pi^2} \oint_C A \left(\frac{\partial \varphi}{\partial n} - \frac{\partial h_1}{\partial n} \right)^2 ds \oint_C \left(x^2 n_x - \frac{\partial h_2}{\partial n} \right)^2 ds. \tag{3.13}$$

We insert the known boundary value for φ on the right-hand side and use the Rayleigh-Ritz technique to obtain close upper and lower bounds for y_s. Note that if the boundary value problem

$$\begin{aligned} \Delta h_2 &= 0 \quad \text{in} \quad \mathscr{D} \\ \frac{\partial h_2}{\partial n} &= x^2 n_x \quad \text{on} \quad C \end{aligned} \tag{3.14}$$

can be solved, it is then possible to determine y_s explicity without solving either the torsion or the flexure problem.

In a similar way it is possible to obtain upper and lower bounds for the coordinate x_s.

4. References

[1] P. Cicala, Atti. accad. sci. Torino **70**, pp. 356–371 (1935).
[2] J. N. Goodier, J. Aeronaut. Sci. **11,** pp. 272–280 (1944).
[3] A. E. H. Love, A treatise on the mathematical theory of elasticity (Cambridge Univ. Press, London, England, 1927).
[4] W. R. Osgood, J. Appl. Math. [A] **10,** pp. 62–64 (1943).
[5] L. E. Payne and H. F. Weinberger, J. Math. Phys. **4,** pp. 291–307 (1955).
[6] L. E. Payne and H. F. Weinberger, Pacific J. Math. **8,** pp. 551–573 (1958).
[7] C. E. Pearson, Z. angew. Math. u. Mech. **36,** pp. 94–96 (1956).
[8] F. Rellich, Math. Z. **46,** pp. 635–646 (1940).
[9] I. S. Sokolnikoff, Mathematical theory of elasticity, 2d ed. (McGraw-Hill Book Co., Inc., New York, N.Y., 1956).
[10] E. Trefftz, Z. angew. Math. u. Mech. **15,** pp. 220–225 (1935).
[11] A. Weinstein, Quart. Appl. Math. **5,** pp. 97–99 (1947).

Washington, D.C.

(Paper 64B2–28)

JOURNAL OF RESEARCH of the National Bureau of Standards—B. Mathematics and Mathematical Physics
Vol. 64B, No. 2, April-June 1960

Half-Round Inductive Obstacles in Rectangular Waveguide

D. M. Kerns

(January 30, 1960)

Formulas are derived for the accurate calculation of the lowest-mode, lumped-element representation of perfectly conducting half-round inductive obstacles in rectangular waveguide. These obstacles consist of either one or two opposed semicircular cylindrical indentations extending across the narrow sides of the waveguide. They seem especially suitable for use as precise calculable standards of reflection or impedance in waveguide. Schwinger's integral equation approach [1,2] [1] is used to obtain stationary expressions for the desired parameters as functionals of the surface currents on the obstacles. Upper bounds are obtained for one of the two parameters. Explicit formulas are derived for the values of the parameters under the assumption of n-term Fourier sine-series expansions for the obstacle currents. Rapid convergence is indicated by numerical evaluations for $n=1$, 2, and 3. In the process of obtaining expressions suitable for numerical calculation, an expansion (believed to be new) of the Green's function of the problem is obtained and the sums of certain infinite series of Bessel's functions occurring in this expansion are expressed in terms of definite integrals. A brief numerical table of these sums, sufficient for the evaluation of the $n=1$ approximation, is included.

1. Introduction

In this paper formulas are derived for the accurate calculation of lowest-mode, lumped-element parameters for what may be identified as "single half-round" and "double half-round" inductive obstacles in rectangular waveguide. As shown in figure 1, the obstacles consist of semicircular cylindrical indentations extending across the narrow sides of the waveguides.

The particular geometry considered seems especially suitable for obstacles to be used as calculable standards of reflection or impedance in waveguide. The geometry is well suited to electroforming, so that obstacles may be fabricated by this process as well as by machining. For obstacles producing standing-wave ratios of moderate values, the obstacle radius is large compared to high-standard machining tolerances, so that unduly close tolerances are not required. A change in waveguide cross section before and after the obstacle is avoided, eliminating the need for waveguide components in odd sizes and permitting maximum flexibility of interconnection.

In the present work the obstacle and waveguide surfaces are assumed perfectly conducting.

The *Waveguide Handbook* [3] contains approximate formulas for semi-elliptical obstacles,[2] which can be specialized to apply to the present problem. However, for the contemplated use in standards work, formulas with very much greater accuracy (10^3 or 10^4 times greater) are wanted.

Brief descriptions of the present and some related work have been given previously [4].

2. Formulation of Problem

We employ the customary complex electric and magnetic field vectors $\boldsymbol{E}=\boldsymbol{E}(\boldsymbol{r})$, $\boldsymbol{H}=\boldsymbol{H}(\boldsymbol{r})$, which satisfy Maxwell's equations in the form

$$\left.\begin{aligned}\nabla\times\boldsymbol{E}&=-j\omega\mu\boldsymbol{H},\\ \nabla\times\boldsymbol{H}&=j\omega\epsilon\boldsymbol{E}+\boldsymbol{J},\end{aligned}\right\} \tag{2.1}$$

[1] Figures in brackets indicate the literature references at the end of this paper.

[2] The pertinent formulas are afflicted with a number of misprints; correct forms of these results for half-round obstacles are given later in this paper.

under the assumption of time-dependence represented by the (omitted) factor exp ($j\omega t$). Here t is the time, $\omega/(2\pi)$ is the frequency, j is the imaginary unit, $\boldsymbol{J}$ is the complex current-density vector, and μ and ϵ are respectively the permeability and the permittivity of the homogeneous, isotropic, nondissipative medium in the waveguide. The MKS system of units is employed.

As shown in figure 1, we choose a rectangular coordinate system $Oxyz$ such that the interior of the waveguide is the space $0<x<a$, $0<y<b$, $-\infty<z<\infty$ and the surfaces of the obstacles are given by the loci

$$\left.\begin{aligned} R^2&=x^2+z^2, \quad 0\leqq x\leqq R, \\ R^2&=(a-x)^2+z^2, \quad 0\leqq a-x\leqq R \end{aligned}\right\} \tag{2.2}$$

in the double half-round case and by

$$R^2=x^2+z^2, \qquad 0\leqq x\leqq R \tag{2.3}$$

in the single half-round case.

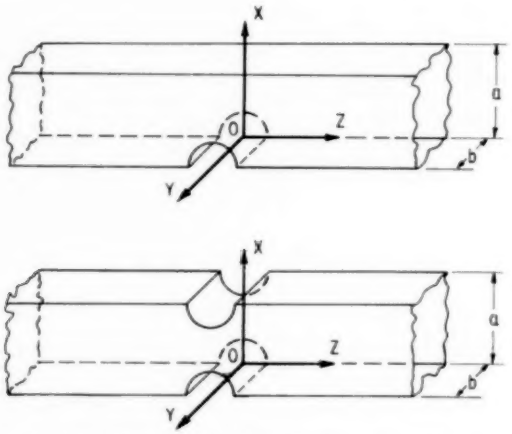


FIGURE 1. *Single and double half-round inductive obstacles in rectangular waveguide.*

The fields incident on the obstacles are, by hypothesis, to be in the TE_{10} mode. Traveling-wave forms for this mode may be written

$$\left.\begin{aligned} \boldsymbol{E}^{\pm}&=\sin\ (\pi x/a)e^{\mp j\beta z}\boldsymbol{e}_y, \\ \boldsymbol{H}^{\pm}&=[\mp\eta\sin\ (\pi x/a)\boldsymbol{e}_x+j\pi(\omega\mu a)^{-1}\cos\ (\pi x/a)\boldsymbol{e}_z]e^{\mp j\beta z}. \end{aligned}\right\} \tag{2.4}$$

The upper and lower signs refer respectively to waves going in the positive and the negative z-directions; the phase constant β is equal to $[k^2-(\pi/a)^2]^{1/2}$, where $k=\omega(\mu\epsilon)^{1/2}$; the wave-admittance $\eta=\beta/(\omega\mu)$; and $\boldsymbol{e}_x$, $\boldsymbol{e}_y$, $\boldsymbol{e}_z$ are the unit vectors of the system $Oxyz$. It is assumed that $\pi<ka<2\pi$, so that the TE_{10} mode, but no higher mode of the type TE_{n0}, will have real propagation.

We define "voltage" and "current" $v_m(z_m)$,$i_m(z_m)$ for the TE_{10} mode by means of the equations

$$\left.\begin{aligned} E_{my}&=v_m(z_m)\sin\ (\pi x/a), \\ H_{mx}&=(-)^m i_m(z_m)\eta\sin\ (\pi x/a). \end{aligned}\right\} m=1,2 \tag{2.6}$$

Here the indices $m=1,2$ refer respectively to the "left-hand" side ($z_1<-R$) and the "right-hand" ($z_2>R$) of the obstacle considered, and E_{my}, H_{mx} are components (identified by the subscripts) of the TE_{10}-mode part of whatever total electromagnetic field may be present in the waveguide at $z=z_m$. From (2.6) and (2.4) it follows that the waveguide characteristic impedance, defined as the value of $v_m(z_m)/i_m(z_m)$ for a pure traveling wave incident on the side m of the obstacle, is equal to unity for $m=1,2$.

The desired lowest-mode, lumped-element description of an obstacle is contained in its impedance matrix Z, which characterizes the linear relations

$$v_1 = Z_{11} i_1 + Z_{12} i_2, \qquad v_2 = Z_{21} i_1 + Z_{22} i_2 \tag{2.7}$$

imposed by the obstacle on the terminal variables v_m, i_m. The values of the Z_{mn} depend upon the location of the terminal planes $z = z_1$ and $z = z_2$; in the present problem it turns out to be convenient to extrapolate these planes to $z_1 = z_2 = 0$.

Now, the equality $Z_{11} = Z_{22}$ follows from the structural symmetry (and the symmetrical disposition of the terminal planes) with respect to the plane $z = 0$; $Z_{12} = Z_{21}$ is assured by both symmetry and reciprocity; and, since losslessness has been assumed, Z is pure imaginary. Thus Z is of the form

$$\begin{pmatrix} jX_{11} & jX_{12} \\ jX_{12} & jX_{11} \end{pmatrix}, \tag{2.8}$$

and there are only two independent parameters to be determined.

We do not obtain formulas for X_{11} and X_{12} directly; but rather, to exploit the symmetry of the problem more fully, we consider the impedances obtaining under modes of excitation in which the electromagnetic field is either symmetric or antisymmetric with respect to the plane $z = 0$. Thus in the symmetric case, $v_1(0) = v_2(0)$, $i_1(0) = i_2(0)$, and the impedance may be written

$$Z_{ee} = v_{1,e}(0) / i_{1,e}(0), \tag{2.9a}$$

where we have added a subscript distinguishing the "even" case. In the antisymmetric ("odd") case, $v_1(0) = -v_2(0)$, $i_1(0) = -i_2(0)$, and the impedance may be written

$$Z_{00} = v_{1,0}(0) / i_{1,0}(0), \tag{2.9b}$$

again adding a distinguishing subscript. The relations

$$2jX_{11} = Z_{ee} + Z_{00}, \qquad 2jX_{12} = Z_{ee} - Z_{00} \tag{2.10}$$

follow directly from (2.7), (2.8), and (2.9). (It may be noted that the symmetric and antisymmetric field distributions correspond to the eigenvectors of the matrix (2.8) and the quantities Z_{ee} and Z_{00} are the eigenvalues of this matrix.)

An equivalent network for the obstacles is of interest and is presented here for convenient reference. For a T-network, using the sign conventions shown in figure 2, we find

$$Z_1 = Z_{00}, \qquad 2Z_2 = Z_{ee} - Z_{00}, \tag{2.11}$$

where Z_1 and Z_2 are respectively the series and the shunt elements of the symmetrical T. It may be remarked that the obstacles considered are called "inductive" because for small radii the important element, Z_2, is a positive reactance.

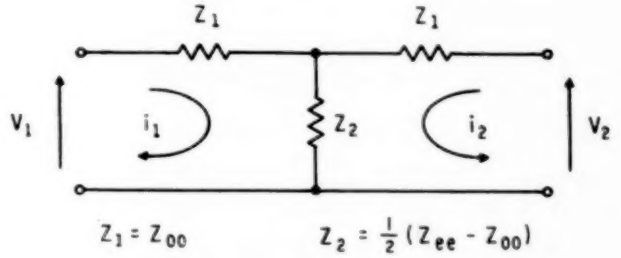


FIGURE 2. *Equivalent network.*

3. Analysis

3.1. Reduction to Scalar Form

Maxwell's equations directly imply the general differential equation for $\boldsymbol{E}$,

$$\nabla\times\nabla\times\boldsymbol{E}=k^2\boldsymbol{E}-j\omega\mu\boldsymbol{J}, \tag{3.1}$$

which must be specialized appropriately for the present problem. $\boldsymbol{E}$ is moreover subject to the boundary condition that its tangential components vanish on perfectly conducting surfaces.

In the present problem both the exciting field and the waveguide-obstacle structure are independent of y in the range $0\leqq y\leqq b$, and it follows that the whole field, secondary as well as primary, will be independent of y in the same range. Further, since the x- and z-components of $\boldsymbol{E}$ must vanish in particular on the $y=0$ surface of the waveguide, these components must vanish for all values of y involved. Thus the electric field is of the form $\boldsymbol{E}=\phi(x, z)\boldsymbol{e}_y$; eq (3.1) (with $\boldsymbol{J}\equiv 0$) reduces to

$$\nabla^2\phi+k^2\phi=0 \qquad \text{in} \quad S,$$

and the boundary condition becomes

$$\phi=0 \qquad \text{on} \quad C,$$

where S is the cross section and C is the boundary of the structure in a plane $y=$ constant (fig. 3).

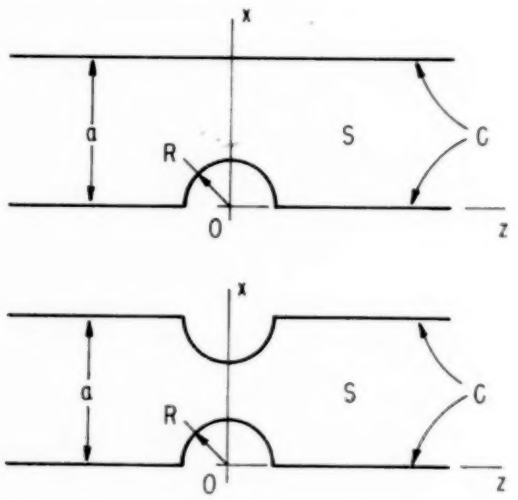


FIGURE 3. *Relevant geometry and coordinates for the single and double half-round problems.*

The magnetic field corresponding to $\boldsymbol{E}$ of the above form is

$$\boldsymbol{H}=(j\omega\mu)^{-1}\left(\frac{\partial\phi}{\partial z}\boldsymbol{e}_x-\frac{\partial\phi}{\partial x}\boldsymbol{e}_z\right). \tag{3.2a}$$

The surface current $\boldsymbol{K}=\boldsymbol{H}\times\boldsymbol{n}$ on C is thus given by

$$\boldsymbol{K}=-(j\omega\mu)^{-1}\boldsymbol{n}\cdot\nabla\phi\boldsymbol{e}_y \tag{3.2b}$$

evaluated on C, where $\boldsymbol{n}$ is the outward normal unit vector on C. Thus $\boldsymbol{K}$ has only a y-component and may be written as $K(x,z)\boldsymbol{e}_y$.

Inasmuch as the whole problem can be dealt with in terms of ϕ and K, the problem is essentially a two-dimensional scalar one. In what follows it will be convenient to refer to ϕ simply as "the field".

3.2. Integral Equation Formulation

We now introduce a Green's function $\Gamma=\Gamma(x, z, x', z')$ such that Γ is the field that would be produced in the unperturbed waveguide at the point (x, z) by a unit current filament parallel to $\boldsymbol{e}_y$ at the point (x', z'). As a function of (x, z), Γ satisfies the differential equation

$$\nabla^2\Gamma+k^2\Gamma=j\omega\mu\ \delta(x-x')\ \delta(z-z') \tag{3.3}$$

obtained from (3.1) with $\boldsymbol{J}=\delta(x-x')\ \delta(z-z')\boldsymbol{e}_y$, where δ denotes the Dirac delta-function. Γ further satisfies the boundary condition $\Gamma=0$ for $x=0,\ a$ and represents outgoing waves for $|z-z'|\to\infty$. An explicit form for Γ is [5]

$$\Gamma(x, z, x', z')=-\omega\mu(\beta a)^{-1}\sin(\pi x/a)\sin(\pi x'/a)e^{-j\beta|z-z'|}$$
$$-(j\omega\mu/a)\sum_{n=2}^{\infty}\alpha_n^{-1}\sin(n\pi x/a)\sin(n\pi x'/a)e^{-\alpha_n|z-z'|}, \tag{3.4}$$

where $\alpha_n=\sqrt{(n\pi/a)^2-k^2}$ for $n\geq 2$ and the other symbols are as already defined.

It is convenient to put

$$\Gamma=-\omega\mu(F+jG), \tag{3.5}$$

where F and G are real. We observe that F can be written in the form

$$F=(\beta a)^{-1}[\psi_e(x, z)\ \psi_e(x', z')+\psi_0(x, z)\ \psi_0(x', z')], \tag{3.6a}$$

with the definitions

$$\left.\begin{aligned}\psi_e(x, z)&=\cos\beta z\ \sin(\pi x/a),\\ \psi_0(x, z)&=\sin\beta z\ \sin(\pi x/a).\end{aligned}\right\} \tag{3.6b}$$

The functions ψ_e and ψ_0 are respectively symmetric and antisymmetric with respect to reflection in the plane $z=0$.

With the aid of the Green's function we can set up the fundamental equations of the problem in integral-equation form. Let $\psi=\psi(x, z)$ denote the exciting field—i.e., the field, due to remote sources, that would exist in the absence of obstacles. Then, with an obstacle present, the total field ϕ can be written as the sum of the exciting field and the secondary field due to the current on the obstacle:

$$\phi(x,z)=\psi(x,z)+\int_{C_0}\Gamma(x,z,x',z')K(x',z')\,dC', \tag{3.7}$$

where K is the surface current on the obstacle and the integral is a line integral going over the contour C_0 of the obstacle (C_0 consists of two parts in the case of the double half-rounds). If the point (x, z) is taken on C_0, ϕ must vanish and we have

$$0=\psi(x,z)+\int_{C_0}\Gamma(x,z,x',z')K(x',z')\,dC', \qquad (x,z)\quad\text{on } C_0. \tag{3.8}$$

This is an inhomogeneous integral equation of the first kind determining K for a given ψ. The equation reduces to slightly different forms in the symmetric and antisymmetric cases, which we now consider.

a. Symmetric Case

Let the exciting field be $A\psi_e(x, z)$, where A is an arbitrary amplitude and ψ_e, as defined in (3.6), is a symmetric function of z. The surface current on the obstacle will then also be a symmetric function of z; we denote it by K_e. We now examine (3.7) for $z<<z'$, say, to find the lowest-mode component and thus to find $r_{1,e}(z)$ and $i_{1,e}(z)$. For $z<<z'$, (3.7) becomes

$$\phi(x,z)=A\cos\beta z\ \sin(\pi x/a)-(\beta a)^{-1}\omega\mu\ \sin(\pi x/a)e^{j\beta z}\int_{C_0}\psi_e K_e\,dC, \tag{3.9}$$

since integrals of the type $\int_{C_0} \psi_0 K_e \, dC$ vanish because of the mutual orthogonality of symmetric and antisymmetric functions. By comparison with (2.6) it is seen that

$$v_{1,e}(z) = A \cos \beta z - (\beta a)^{-1} \omega\mu \, e^{j\beta z} \int_{C_0} \psi_e K_e \, dC. \tag{3.10}$$

Using (3.2a) to obtain the x-component of the magnetic field associated with (3.9) and again referring to (2.6) one finds

$$i_{1,e}(z) = -jA \sin \beta z + (\beta a)^{-1} \omega\mu \, e^{j\beta z} \int_{C_0} \psi_e K_e \, dC. \tag{3.11a}$$

Equations (3.10) and (3.11a) hold directly from the definitions for $z < -R$ and by extrapolation for $z = 0$. In particular (3.11a) yields

$$i_{1,e}(0) = (\beta a)^{-1} \, \omega\mu \int_{C_0} \psi_e K_e \, dC. \tag{3.11b}$$

Next, in the integral eq (3.8), we separate the Green's function into real and imaginary parts as in (3.5), use the orthogonality property of even and odd functions, and thus find

$$\psi_e(x, z) \, v_{1,e}(0) = j\omega\mu \int_{C_0} G(x, z, x' z') \, K_e(x', z') \, dC'. \tag{3.12}$$

This equation and (3.11b) together furnish a definitive mathematical statement of the symmetric part of our problem.

b. Antisymmetric Case

In this case we let the exciting field be $A\psi_0(x, z)$, where A is again an arbitrary constant and ψ_0, defined in (3.6), is an antisymmetric function of z. The current, K_0, on the obstacle will then also be antisymmetric with respect to z. In the same manner as in the symmetric case one finds, for $z \gtrless 0$,

$$v_{1,0}(z) = A \sin \beta z + (\beta a)^{-1} j\omega\mu e^{j\beta z} \int_{C_0} \psi_0 K_0 dC. \tag{3.14a}$$

In particular,

$$v_{1,0}(0) = (\beta a)^{-1} j\omega\mu \int_{C_0} \psi_0 K_0 dC. \tag{3.14b}$$

Further,

$$i_{1,0}(z) = jA \cos \beta z - (\beta a)^{-1} j\omega\mu e^{j\beta z} \int_{C_0} \psi_0 K_0 dC. \tag{3.15}$$

and the integral equation becomes

$$\psi_0(x, z) i_{1,0}(0) = -\omega\mu \int_{C_0} G(x, z, x', z') K_0(x', z') dC'. \tag{3.16}$$

This equation and (3.14b) together furnish a definitive statement of the antisymmetric cases of our problem.

3.3. Summary of all Cases

In the case of the double half-rounds the above integrations go over a two-part contour consisting of an upper semicircle C_u and a lower semicircle C_l. An equivalent problem involving only one of the parts of the contours, say C_l, can be formulated using the fact that the exciting field as well as the structure is symmetric with respect to the plane $x = a/2$, so that the current on the obstacles must also be symmetric with respect to this plane. This symmetry is quite

independent of the symmetries with respect to $z=0$ distinguished by the subscripts "e" and "o," and so these subscripts will be dropped for the moment. Thus, it is clear that

$$\int_{C_o} \psi K\, dC = \int_{C_u} \psi K\, dC + \int_{C_l} \psi K\, dC = 2\int_{C_l} \psi K\, dC. \tag{3.18}$$

Further,

$$\int_{C_o} G(x,z,x',z')\, K(x',z')\, dC' = \int_{C_l} G^{(2)}(x,z,x',z')\, K(x',z')\, dC', \tag{3.19}$$

where $G^{(2)}$ is defined by

$$G^{(2)}(x,z,x',z') = G(x,z,x',z') + G(x,z,a-x',z').$$

Equations (3.18) and (3.19) enable the desired restatement of the double half-round problems.

In what follows, superscripts 1,2 will be used when it is desired to distinguish quantities associated with the single and the double half-round problems, respectively (this means in particular $G \equiv G^{(1)}$). Integrals will be indicated by means of the convenient scalar product notation, e.g.,

$$(\psi, K) = \int_{C_l} \psi K\, dC, \tag{3.20a}$$

and by the operator notation

$$GK = \int_{C_l} G(x,z,x',z')\, K(x',z')\, dC'. \tag{3.20b}$$

As indicated, all such integrals are to be taken over C_l, the lower semicircle. As a further notational convenience, we introduce the "normalized" surface currents,

$$I_e = j\omega\mu K_e / v_{1,e}(0), \qquad I_o = -\omega\mu K_o / i_{1,o}(0), \tag{3.21}$$

for both the single and the double problems.

Our results thus far may now be summarized as follows. For the symmetric cases we have

$$\psi_e = G^{(h)} I_e, \qquad 1/X_{ee}^{(h)} = h(\psi_e, I_e)/(\beta a) \tag{3.22}$$

and for the antisymmetric cases

$$\psi_o = G^{(h)} I_o, \qquad X_{oo}^{(h)} = -h(\psi_o, I_o)/(\beta a), \tag{3.23}$$

where $h=1$ or 2 for the single or the double half-rounds, respectively, and X_{ee}, X_{oo} are the reactances corresponding to the (pure imaginary) impedances Z_{ee}, Z_{oo} defined in (2.9).

4. Solution for the Reactances

4.1. Application of Rayleigh-Ritz Method

In each of the problems specified by (3.22) and (3.23), an unknown reactance or susceptance is proportional to a scalar product

$$M = (\psi, I), \tag{4.1a}$$

where the function I is determined by an integral equation,

$$GI = \psi, \tag{4.1b}$$

in which ψ and G are given and G is symmetric. From these equations one may easily construct the "stationary representation" [1, 2]

$$M = (\psi, I)^2/(I, GI) \tag{4.2}$$

for M. This constitutes the starting point for the method of Rayleigh and Ritz [6], which we wish to use to obtain an approximate expression for M. One assumes in (4.2) an approximation for the surface current in the form

$$I_n = x_i f_i, \tag{4.3}$$

where the x_i are coefficients to be determined, the f_i are members of a suitable set of basis functions (to be chosen explicitly), and summation from 1 to n over repeated indices $i, j, \ldots$, is understood. Expression (4.2) becomes an ordinary function of the x_i and is to be subjected to the conditions of stationarity $\partial M/\partial x_i = 0$, $i = 1, 2, \ldots, n$. This leads to the system of equations

$$G_{ij}x_j = (x_k c_k / M_n)c_i$$

where

$$c_i \equiv (\psi, f_i), \qquad G_{ij} \equiv (f_i, Gf_j), \tag{4.4}$$

and M_n denotes the now-determinate approximate value for M. If the arbitrary normalization of the x_i is chosen so that $x_k c_k = M_n$, then

$$x_i = (G^{-1})_{ij} c_j,$$

where $(G^{-1})_{ij}$ denotes an element of the inverse of the $n \times n$ matrix of the G_{ij}, and it immediately follows that

$$M_n = (G^{-1})_{ij} c_i c_j.$$

Finally, we write this quadratic form as a ratio of determinants,

$$-M_n = \frac{\begin{vmatrix} 0 & c_1 & \cdots & c_n \\ c_1 & G_{11} & \cdots & G_{1n} \\ \cdots & \cdots & \cdots & \cdots \\ c_n & G_{n1} & \cdots & G_{nn} \end{vmatrix}}{\begin{vmatrix} G_{11} & \cdots & G_{1n} \\ \cdots & \cdots & \cdots \\ G_{n1} & \cdots & G_{nn} \end{vmatrix}}. \tag{4.5}$$

Thus M_n is expressed in terms of the known quantities (4.4); if the objective is only to calculate the reactances, it is not necessary to calculate the x_i.

4.2. Definiteness of G; Upper Bounds for X_{00}

Although in our problem G is not a definite kernel, it does have definite character with respect to a suitably restricted class of admissible functions. In fact, G is positive definite with respect to functions that are antisymmetric with respect to z and vanish for $|z| > \lambda_g/4$, where $\lambda_g = 2\pi/\beta$ is the "guide wavelength" of the single propagating mode.[3] In terms of the integrals of interest here, this means that if $\tilde{I}_0$ is any assumed surface current distribution, continuous, not identically equal to zero, and antisymmetric with respect to z, then

$$(\tilde{I}_0, G\tilde{I}_0) > 0, \tag{4.6}$$

provided $R < \lambda_g/4$.

Now, for a positive definite kernel it is well known that the right-hand side of (4.2) is not only stationary, but also a maximum for the true I. For this result we offer the following brief proof based directly on the Schwarz inequality.[4] Assuming G to be positive definite, we wish to compare the approximate value for M given by

$$\tilde{M} = (\tilde{I}, \psi)^2/(\tilde{I}, G\tilde{I}),$$

[3] This result is suggested by qualitative physical considerations and has been verified analytically by the author. It has also been verified that G is not definite with respect to functions symmetric with respect to z.

[4] Proofs of this result in the literature known to the author, e.g., [1,2], employ the so-called bilinear expansion for the kernel involved and do not yield the "only if" part of the result (4.7).

where $\tilde{I}$ is an admissible approximation for the surface current, with the true value given by (4.1a) (as well as by (4.2)) when $\tilde{I}$ satisfies (4.1b). From (4.1b) it follows that $(\tilde{I}, \psi)=(\tilde{I}, GI)$; the Schwarz inequality for positive-definite symmetric transformations [7] gives

$$(\tilde{I}, GI)^2 \leqq (\tilde{I}, G\tilde{I})\ (I, GI).$$

Since $(I, GI)=(I,\psi)=M$, we have the result

$$\tilde{M} \leqq M. \tag{4.7}$$

Moreover the sign of equality holds in the Schwarz inequality and hence in this result if and only if $\tilde{I}$ is proportional to I.

Clearly, (4.7) is applicable when G is merely positive definite with respect to all functions that need be admitted; by (4.6) this will be the case in the antisymmetric parts of our problem provided $R<\lambda_g/4$. Hence, subject to this inequality, in the antisymmetric cases we shall have $M_n<M$ (we do not anticipate an exact result for finite n); this in turn implies that numerical values for X_{00} will be (algebraic) upper bounds for the true values—assuming, of course, that numerical evaluations involved are sufficiently precise.

Unfortunately one can not obtain a bound for X_{ee} in this simple manner. Furthermore, derived quantities of practical interest (such as (4.17)) will usually depend upon both X_{ee} and X_{00}. Thus in general no bounds for such derived quantities are determined.

4.3. Basis Functions

We introduce plane polar coordinates r,θ such that

$$x=r \sin \theta, \qquad z=r \cos \theta. \tag{4.8}$$

The path of integration C_l is then given parametrically by

$$x=R \sin \theta, \qquad z=R \cos \theta, \tag{4.9}$$

where R is the obstacle radius and $0 \leqq \theta \leqq \pi$. On this path the surface current $I=I(\mathrm{R} \sin \theta, R \cos \theta)$ becomes a function of θ, which we denote simply by $I(\theta)$. We observe that symmetries with respect to $z=0$ are equivalent to symmetries with respect to $\theta=\pi/2$.

As basis functions we take

$$f_i=2(\pi R)^{-1} \sin (2i-1)\theta \tag{4.10}$$

for the symmetric cases (expansion of I_e), and

$$f_i=2(\pi R)^{-1} \sin 2i\theta \tag{4.11}$$

for the antisymmetric cases (expansion of I_0). The two sets of functions are complete for the expansion of symmetric and antisymmetric functions, respectively, in the interval $(0,\pi)$. The completeness gives good assurance that for sufficiently large n, M_n will approach arbitrarily close to the true M. For practical numerical calculation it is vital that the convergence of M_n be rapid. In the present instance rapid convergence might be anticipated on the grounds that the true I_e and I_0 may be expected to be smooth (possibly infinitely smooth) functions of θ vanishing at 0 and π, so that their Fourier sine series should converge rapidly.[5]

---

[5] The points $(R, 0)$ and (R, π) are the vertices of the corners formed where the half-round centered at O meets the waveguide wall. The vanishing of the surface current in these vertices is equivalent, by (3.2), to the vanishing of the components of the surface magnetic field normal to the vertices. That these field components should vanish in the limit as a vertex is approached follows from properties of Sommerfeld's solution [14] of the problem of diffraction by a wedge of arbitrary angle.

4.4. Evaluation of Integrals

Integrals of the type (ψ, f_i) may be obtained relatively easily with the aid of the well-known expansion [8]

$$e^{j\zeta \sin \alpha}=\sum_{n=-\infty}^{\infty} J_n(\zeta)e^{jn\alpha},$$

wherein J_n is the n^{th} order Bessel function of the first kind. In this expansion we put $\zeta=kR$ and $\alpha=\pm\theta-\chi+(\pi/2)$, with $k \sin \chi=\pi/a$, $k \cos \chi=\beta$, and employ (4.9). The expansion

$$e^{j\beta z} \sin (\pi x/a)=2\sum_{n=1}^{\infty} j^{n-1}J_n(kR) \sin n\theta \sin n\chi \tag{4.12}$$

is then easily obtained. The integrations yielding the desired quantities are now elementary; one obtains

$$c_i=2(-)^{(p-1)/2} J_p(kR) \sin p\chi, \qquad p=2i-1, \tag{4.13a}$$

for the symmetric cases, and

$$c_i=2(-)^{(p-2)/2} J_p(kR) \sin p\chi, \qquad p=2i, \tag{4.13b}$$

for the antisymmetric cases. (It is encouraging that these quantities decrease rapidly with increasing index.)

The evaluation of the integrals of the type $(f_i, Gf_j)=G_{ij}$ is considerably more complicated than the evaluation of the c_i, and the task has been relegated to the appendix The results obtained may be summarized conveniently in the following form, which comprehends four cases:

$$\frac{G_{ij}^{(h)}}{c_i c_j}=-\frac{1}{4s_p s_q}\left[\rho_p(kR)\delta_{pq}+\sigma_{pq}^{(h)}(ka)\right] \tag{4.14}$$

Here $s_p \equiv \sin p\chi$; $\rho_p(kR) \equiv Y_p(kR)/J_p(kR)$; in the symmetric cases, $p=2i-1$ and $q=2j-1$; in the antisymmetric cases, $p=2i$ and $q=2j$; the functions $\sigma_{pq}^{(h)}$ are defined in the appendix (and tabulated briefly in table 2); Y_p denotes the Bessel function of the second kind; and all other symbols involved have been defined previously.

4.5. Results

The value of a reactance element corresponding to an n-term expansion (4.3) will be called an "n^{th} approximation" and will be written simply $X_{ee}^{(h)}$ or $X_{00}^{(h)}$ without special notation indicating the value of n being considered. An explicit expression, in determinantal form, for $X_{ee}^{(h)}$ in the n^{th} approximation may be obtained by combining (3.22), (4.5), and (4.14); viz,

$$X_{ee}^{(h)}=\frac{\beta a}{4h}\frac{\begin{vmatrix} \rho_1+\sigma_{11}^{(h)} & \cdots & \sigma_{1,2n-1}^{(h)} \\ \cdots & \cdots & \cdots \\ \sigma_{1,2n-1}^{(h)} & \cdots & \rho_{2n-1}+\sigma_{2n-1,2n-1}^{(h)} \end{vmatrix}}{\begin{vmatrix} 0 & s_1 & \cdots & s_{2n-1} \\ s_1 & \rho_1+\sigma_{11}^{(h)} & \cdots & \sigma_{1,2n-1}^{(h)} \\ \cdots & \cdots & \cdots & \cdots \\ s_{2n-1} & \sigma_{1,2n-1}^{(h)} & \cdots & \rho_{2n-1}+\sigma_{2n-1,2n-1}^{(h)} \end{vmatrix}} \tag{4.15}$$

(In this and in the following equation the arguments kR and ka are understood.) Similarly, (3.23), (4.5), and (4.14) yield

$$X_{00}^{(h)}=-\frac{4h}{\beta a}\frac{\begin{vmatrix} 0 & s_2 & \cdots & s_{2n} \\ s_2 & \rho_2+\sigma_{22}^{(h)} & \cdots & \sigma_{2,2n}^{(h)} \\ \cdots & \cdots & \cdots & \cdots \\ s_{2n} & \sigma_{2,2n}^{(h)} & \cdots & \rho_{2n}^{(h)}+\sigma_{2n,2n}^{(h)} \end{vmatrix}}{\begin{vmatrix} \rho_2+\sigma_{22}^{(h)} & \cdots & \sigma_{2,2n}^{(h)} \\ \cdots & \cdots & \cdots \\ \sigma_{2,2n}^{(h)} & \cdots & \rho_{2n}+\sigma_{2n,2n}^{(h)} \end{vmatrix}}. \tag{4.16}$$

The determinantal form of these expressions seems convenient enough for practical calculations with n at least as large as 3. However, it may be remarked that, because of the special form of the diagonal elements, the determinants can be expanded in the same way as certain determinants occurring in the Fredholm theory of integral equations. This expansion may be convenient for the consideration of arbitrarily large n, but is not needed here.

Once the reactances X_{ee} and X_{00} are determined, other quantities or parameters associated with the representation of an obstacle are of course also determined. In particular, the "voltage" standing-wave ratio seen on one side of an obstacle when the other side is terminated in a matched (reflectionless) load is a familiar quantity having immediate physical significance, and it is convenient to discuss some of the results in terms of this quantity. Recalling that waveguide characteristic impedances were chosen equal to unity and referring to the equivalent network, figure 2, it is easily found that the standing-wave ratio in question is given by

$$\eta=(1+|S|)/(1-|S|), \tag{4.17a}$$

where $|S|$, the magnitude of the associated reflection coefficient, is

$$|S|=\left[1+\left(\frac{X_{ee}-X_{00}}{X_{ee}X_{00}+1}\right)^2\right]^{-1/2}. \tag{4.17b}$$

In order to get some idea of the behavior of the sequence of approximations, the reactance elements and the related VSWR (4.17) have been calculated for $n=1$, 2, and 3 and for a range of values of ka and kR. Table 1 presents more or less typical results in terms of the values of $VSWR$. The convergence of the sequence of approximations, judged on the basis of numerical results exemplified in the table, appears to be very rapid; by the same token, the first approximation furnishes a rather good result for VSWR's up to about 2.

Thus far it has not proved feasible to determine the manner of convergence of the sequence of approximations by theoretical means. In view of the character of the particular problem at hand, it seems probable that the numerical evidence may be relied upon.

TABLE 1. *Convergence of VSWR*

$ka=4.5$

kR	Approximation	VSWR, Single half-round	VSWR, Double half-round
0.2	1	1.0370968	1.0776492
	2	1.0370970	1.0776499
	3	1.0370970	1.0776499
0.7	1	1.4547668	2.8411581
	2	1.4554655	2.8416266
	3	1.4554655	2.8416268
1.0	1	2.1070625	15.900756
	2	2.1125072	16.009131
	3	2.1125112	16.009479

Note: Maximum computational inaccuracy in the above figures is estimated to be approximately ±1 in the sixth significant figure; eight figures, as given, may be significant in indicating the behavior of successive approximations.

Further calculations have been made regarding ka and R/a as independent variables. This choice of variables corresponds to the practical situation where one has a given obstacle (having a given value of R/a) that is to be operated at various frequencies. Figures 4 and 5 show the matched-termination VSWR (4.17) resulting for various values of R/a and f/f_c, where f is the operating frequency and f_c is the cut-off frequency for the TE_{10} mode (and $f/f_c=ka/\pi$). (Precise tables (based on the third approximations) giving basic parameters and possibly some derived quantities are being prepared for separate publication.)

The first approximations appear to be of useful accuracy for some purposes and represent an appreciable extension of previously available results. Consequently, a brief tabulation of the σ-functions, sufficient for the calculation of first approximations, is presented as table 2. For convenience of reference we write down (4.15) and (4.16) for the special case $n=1$,

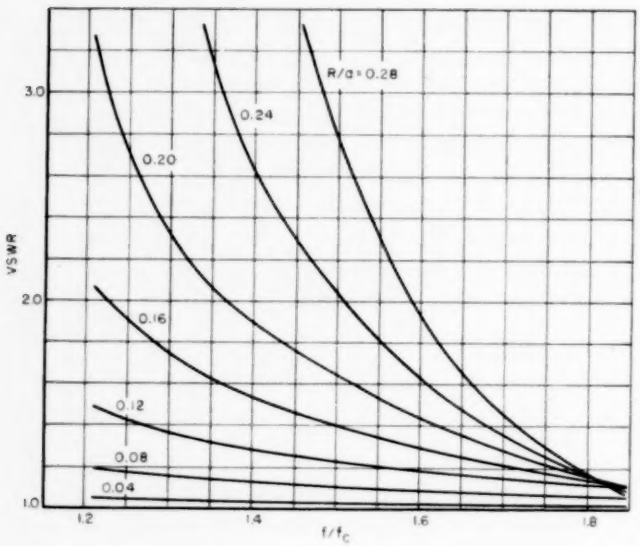


FIGURE 4. *VSWR versus f/f_c for single half-rounds.*

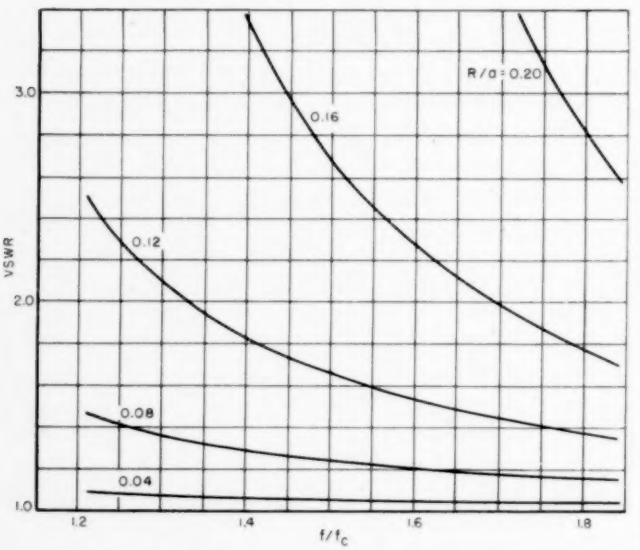


FIGURE 5. *VSWR versus f/f_c for double half-rounds.*

$$X_{ee}^{(h)}=-\frac{\beta a}{4hs_1^2}\left[\rho_1(kR)+\sigma_{11}^{(h)}(ka)\right], \tag{4.18}$$

$$X_{00}^{(h)}=\frac{4hs_2^2}{\beta a}\left[\rho_2(kR)+\sigma_{22}^{(h)}(ka)\right]^{-1}, \tag{4.19}$$

and note that $s_1=\pi/(ka)$ and $s_2=2\pi\beta/(k^2a)$.

TABLE 2. *Values of $\sigma_{pp}^{(h)}(x)$*

x	$\sigma_{11}^{(1)}(x)$	$\sigma_{22}^{(1)}(x)$	$\sigma_{11}^{(2)}(x)$	$\sigma_{22}^{(2)}(x)$
3.80	2.055435	−0.3170852	3.960997	−2.544410
3.85	1.976710	−.1982432	3.857142	−2.215670
3.90	1.899906	−.09078960	3.756554	−1.912736
3.95	1.823995	.006197600	3.659060	−1.633535
4.00	1.749764	.09356160	3.564510	−1.376168
4.05	1.676799	.1720644	3.472757	−1.138910
4.10	1.604988	.2424036	3.383659	−0.9201812
4.15	1.534222	.3052036	3.297087	−.7185580
4.20	1.464400	.3610476	3.212923	−.5327088
4.25	1.395414	.4104564	3.131048	−.3614372
4.30	1.327159	.4539056	3.051353	−.2036444
4.35	1.259533	.4918312	2.973735	−.05831720
4.40	1.192430	.5246272	2.898094	.0746600
4.45	1.125746	.5526568	2.824341	.1985600
4.50	1.059372	.5762488	2.752387	.3117472
4.55	0.9931968	.5956964	2.682146	.4157360
4.60	.9271072	.6112796	2.613542	.5112028
4.65	.8609836	.6232416	2.546500	.5987500
4.70	.7946988	.6318024	2.480946	.6789328
4.75	.7281208	.6371680	2.416812	.7522756
4.80	.6611076	.6395192	2.354034	.8192564
4.85	.5935064	.6390196	2.292551	.8803180
4.90	.5251500	.6358108	2.232301	.9358596
4.95	.4558572	.6300236	2.173228	.9862656
5.00	.3854276	.6217692	2.115279	1.031884
5.05	.3136376	.6111440	2.058400	1.073038
5.10	.2402376	.5982284	2.002545	1.110027
5.15	.1649416	.5830836	1.947661	1.143125
5.20	.08742720	.5657628	1.893705	1.172594
5.25	.007320000	.5462944	1.840631	1.198667
5.30	−.07581400	.5246972	1.788397	1.221571
5.35	−.1624856	.5009664	1.736963	1.241508
5.40	−.2533004	.4750804	1.686288	1.258672
5.45	−.3488884	.4469912	1.636332	1.273236
5.50	−.4504328	.4166292	1.587059	1.285370
5.55	−.4587224	.3838912	1.538434	1.295225
5.60	−.6752200	.3486376	1.490420	1.302940
5.65	−.8016548	.3106864	1.442981	1.308654
5.70	−.9402632	.2697956	1.396088	1.312488
5.75	−1.094004	.2256488	1.349701	1.314555
5.80	−1.266887	.1778340	1.303798	1.314964

Note: Accuracy of the above figures is estimated to be approximately one part per million.

For sufficiently small kR, the expressions just given may be simplified still further. For small kR, the quantities $\rho_1(kR)$ and $\rho_2(kR)$ become large, in accordance with the formula

$$\rho_q(kR) \cong -\frac{1}{\pi} q!\,(q-1)!\left(\frac{2}{kR}\right)^{2q} \qquad (q \gtrsim 1),$$

and thus determine the qualitative behavior of X_{ee} and X_{00}. Using this approximation for the ρ's, neglecting the σ's, and introducing the guide wavelength $\lambda_g = 2\pi/\beta$, one obtains

$$X_{ee}^{(h)} = \frac{2a}{h\lambda_g}\left(\frac{a}{\pi R}\right)^2, \tag{4.20}$$

$$X_{00}^{(h)} = -\frac{ha}{\lambda_g}\left(\frac{\pi R}{a}\right)^4. \tag{4.21}$$

(These are the correct forms of the results in the order of approximation of the more general formulas in the *Waveguide Handbook* [3].) Although these formulas give a rather good picture of the behavior of the reactances, they fail to reflect the interesting qualitative feature that $X_{ee}^{(h)}$ and $X_{00}^{(h)}$ must eventually become equal to each other as R approaches a/h (as can be seen physically).

5. Appendix

5.1. Expansion of Green's Function in Terms of Cylindrical Wave Functions

In order to arrive conveniently at the respective types of Green's functions needed for the single and the double half-round problems, we consider a waveguide of width w (instead of a), bounded by the lines $x=0$ and $x=w$; and, in addition to the electric boundary condition at $x=0$, we consider both electric and magnetic boundary conditions at $x=w$ (by "electric" or "magnetic" boundary conditions it is meant that the Green's function or its normal derivative, respectively, vanishes on the boundary). The radius vectors of the source-point and the field-point will be denoted by $\boldsymbol{r}'$ and $\boldsymbol{r}$, respectively.

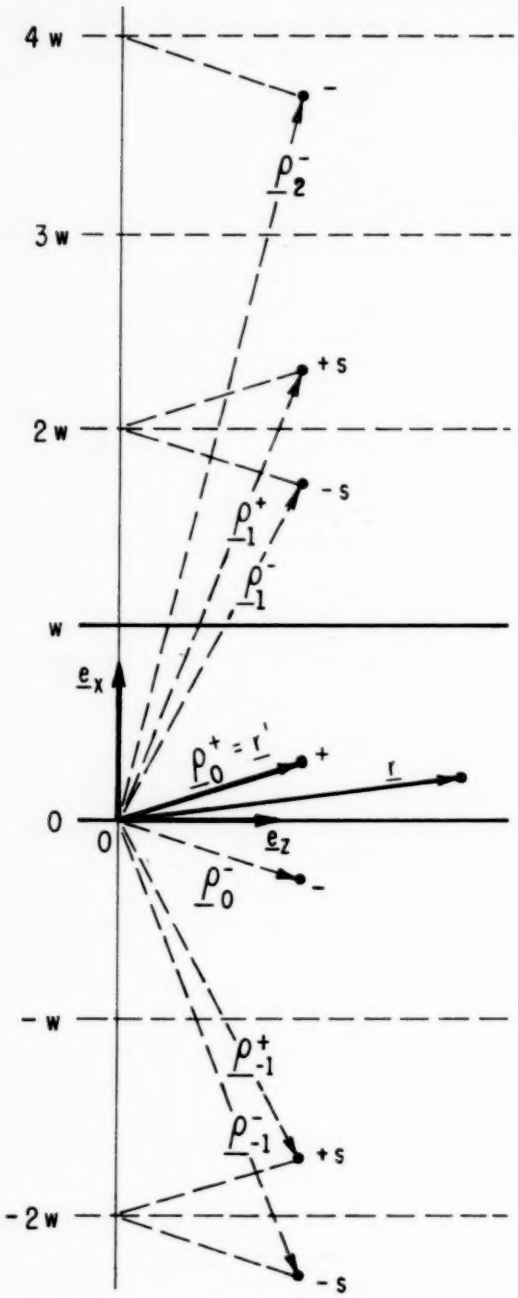
Now, it is well known that the two-dimensional "free-space" Green's function, satisfying (3.3) and representing outgoing waves at infinity, is $-(\omega\mu/4)H_0(k|\boldsymbol{r}'-\boldsymbol{r}|)$, where H_0 denotes the Hankel function of the second kind (as is appropriate for $\exp(j\omega t)$ time dependence). It is also well known that the desired Green's function, Φ, satisfying (3.3) and the required boundary conditions at $x=0$, $x=w$, and at $z=\pm\infty$, can be obtained in the form of a sum of free-space Green's functions by the method of images [9]. In fact

$$\Phi = -(\omega\mu/4)\sum_{n=-\infty}^{\infty} s^n[H_0(k|\boldsymbol{\rho}_n^+ - \boldsymbol{r}|) - H_0(k|\boldsymbol{\rho}_n^- - \boldsymbol{r}|)], \tag{5.1}$$

where $\boldsymbol{\rho}_n^{\pm} = (2nw \pm x')\boldsymbol{e}_x + z'\boldsymbol{e}_z$ and $s=+1$ or -1 according to whether the boundary condition at $x=w$ is electric or magnetic. By inspection of the array of sources (fig. 6) it can be seen that this function has the proper symmetry to satisfy the boundary conditions: it is antisymmetric with respect to $x=0$ and it is antisymmetric or symmetric with respect to $x=w$ according to whether $s=+1$ or -1.

Equation (5.1) can be transformed into a sum of elementary wave-functions of polar coordinates by repeated application of Graf's addition theorem for Bessel functions [10]. We introduce polar coordinates (r,θ), (r',θ'), and (ρ_n,θ_n) for $\boldsymbol{r}$, $\boldsymbol{r}'$, and $\boldsymbol{\rho}_n$ in the manner shown in (4.8), and we assume r and r' both to be less than w, so that $\rho_n > w$ for $n \neq 0$. Considering the triangles with sides $\boldsymbol{\rho}_n$, $\boldsymbol{r}$, and $\boldsymbol{\rho}_n - \boldsymbol{r}$ and applying the addition theorem, we obtain

$$-4(\omega\mu)^{-1}\Phi = \sum_{p=-\infty}^{\infty} \frac{J_p(kr)H_p(kr')}{J_p(kr')H_p(kr)}\, e^{jp\theta}(e^{-jp\theta'} - e^{jp\theta'})$$
$$+ \sum_{n=-\infty}^{\infty}{}' s^n \sum_{p=-\infty}^{\infty} J_p(kr)e^{jp\theta}[H_p(k\rho_n^+)e^{-jp\theta_n^+} - H_p(k\rho_n^-)e^{-jp\theta_n^-}],$$



FIGUEE 6. *Source points in the method of images.*

Relative signs of sources are indicated by the adjacent symbols +, −, +s, and −s; the role of s is defined in connection with (5.1).

where the terms arising from $n=0$ appear in the first line and are to be omitted from the primed summation in the second line; the upper and lower alternatives in the first line apply according as $r<r'$ or $r>r'$, respectively. Again, by the addition theorem,

$$H_p(k\rho_n^{\pm})e^{-jp\theta_n^{\pm}} = \sum_{q=-\infty}^{\infty} J_q(kr')H_{p+q}(|2nkw|)\exp\left[j(\pi/2)\,(q-p)\,\operatorname{sgn} n \pm jq\theta'\right].$$

Combining the last two equations and rearranging terms, one finds that (5.1) can be written in the form

$$\Phi=-\omega\mu\sum_{p=1}^{\infty}\sum_{q=1}^{\infty}\sin p\theta\sin q\theta'\left\{\begin{matrix}J_p(kr)H_p(kr')\\J_p(kr')H_p(kr)\end{matrix}\right\}\delta_{pq}$$
$$+2J_p(kr)J_q(kr')\cos[(\pi/2)(p-q)]\sum_{n=1}^{\infty}s^n[H_{p-q}(2nkw)-H_{p+q}(2nkw)]\Big\}. \quad (5.2)$$

If now one puts $w=a$ and $s=+1$, this expression becomes the desired representation of Γ of the text; similarly, if $w=a/2$ and $s=-1$, (5.2) becomes the desired representation of a Green's function $\Gamma^{(2)}$ related to $G^{(2)}$ in the same way as Γ is related to G (cf. 3.5). Thus, employing the relation (3.5), the relation $H_p=J_p-jY_p$, the notation (3.20), the basis functions (4.10, 4.11), and carrying out the integrations with respect to θ, θ', one obtains

$$\left(\frac{2}{\pi R}\right)^2(\sin p\theta, G^{(h)}\sin q\theta')=-J_p(kR)Y_p(kR)\delta_{pq}-J_p(kR)J_q(kR)\cos[(\pi/2)(p-q)]\sigma_{pq}^{(h)}(ka), \quad (5.3)$$

where the functions $\sigma_{pq}^{(h)}$ are defined as

$$\sigma_{pq}^{(1)}(ka)=2\sum_{n=1}^{\infty}[Y_{p-q}(2nka)-Y_{p+q}(2nka)], \quad (5.4a)$$

$$\sigma_{pq}^{(2)}(ka)=2\sum_{n=1}^{\infty}(-)^n[Y_{p-q}(nka)-Y_{p+q}(nka)]. \quad (5.4b)$$

We note that the above expressions are needed only for positive integer indices p, q that are both even or both odd. We observe also that the form of the expressions is such that those with unequal subscript indices can be evaluated simply in terms of those with equal subscript indices. Further, in view of the easily-verified relation

$$\sigma_{pq}^{(2)}(ka)=2\sigma_{pq}^{(1)}(ka)-\sigma_{pq}^{(1)}(ka/2), \quad (5.5)$$

values or formulas for the σ's "of the second kind" can be derived readily from the same for the σ's "of the first kind." Thus, the evaluation of the $G_{pq}^{(h)}$, which, on the face of it, originally involved double integrals with ka, kR, p, q, and h as independent parameters, has been reduced essentially to the evaluation of sums of the form $\sum_{n=1}^{\infty}Y_{2s}(nx)$ with $\pi<x<2\pi$ or $2\pi<x<4\pi$ and $s=0, 1, 2, \ldots$. It should be observed, however, that these sums converge extremely slowly: for the accuracy desired in the present circumstances the number of terms required would be of the order of 10^{10} or more, at least for the most important σ's. Needless to say, some method of evaluation other than direct summation is indicated.

5.2. Evaluation of the Bessel-Function Series

Two methods of evaluating the series defined in (5.4) will be discussed briefly here.

a. Conversion to Definite Integrals

An evaluation of the desired series may be obtained in terms of definite integrals as follows. We start with the series of Hankel functions (of the second kind)

$$S_\lambda(x)=\sum_{n=1}^{\infty}H_\lambda(nx), \quad (5.6)$$

where x is real, positive, and not an integer multiple of 2π. For the Hankel functions we use the integral representation [11]

$$H_\lambda(x)=\frac{1}{\pi}\int_C e^{-jx\sin\zeta+j\lambda\zeta}d\zeta$$

where $\zeta=\xi+j\eta$ and C is a suitable path, to be chosen in a moment, between the limits $-j\infty$ and $\pi+j\infty$. With the aid of the well-known expression for the sum of a finite number of terms of a geometric series, one finds that the M-term partial sum of (5.6) can be written

$$\sum_{n=1}^{M} H_\lambda(nx)=\frac{1}{\pi}\int_C \frac{e^{j\lambda\zeta}d\zeta}{e^{jx\sin\zeta}-1}-\frac{1}{\pi}\int_C \frac{e^{-jMx\sin\zeta}}{e^{jx\sin\zeta}-1}\,e^{j\lambda\zeta}d\zeta. \tag{5.7}$$

Among the infinitely many poles of the integrands in these integrals, we note in particular ξ_m' and $\xi_m''=\pi-\xi_m'$, where ξ_m' is the principal value of $\sin^{-1}(2m\pi/x)$, $m=0, 1, 2, \ldots, Q$, and Q is the largest integer such that $2Q\pi/x<1$. The path C is now chosen to comprise (except for indentations) the negative imaginary axis, the real axis between 0 and π, and the upper half of the line $\pi+j\eta$; indentations, of radius ρ, are made at each of the above identified poles in the manner shown in figure 7, which illustrates C for the case $Q=1$. Taken on this path, the second integral in (5.7) approaches 0 as M increases indefinitely; as a detailed examination shows, this happens because on the various parts of the path the integrand approaches 0 or it becomes increasingly oscillatory (or both) as $M\to\infty$. Hence we have

$$S_\lambda(x)=\frac{1}{\pi}\int_C \frac{e^{j\lambda\zeta}d\zeta}{e^{jx\sin\zeta}-1}.$$

This result can be expressed in terms of real integrations plus contributions from the indentations. In the limit as $\rho\to 0$, the indentations yield

$$R_o'=\frac{1}{2x}, \qquad R_o''=\frac{e^{j\lambda\pi}}{2x}, \tag{5.8a}$$

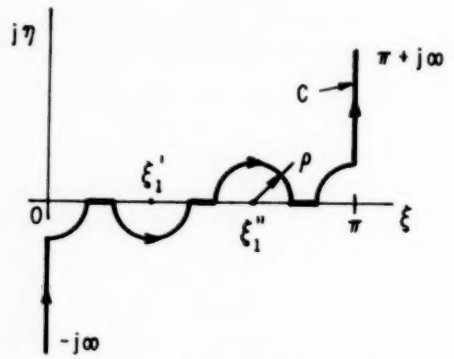


FIGURE 7. *Path of integration.*

and for $m\geq 1$,

$$R_m'=\frac{e^{j\lambda\xi_m'}}{\sqrt{x^2-(2m\pi)^2}}, \qquad R_m''=\frac{e^{j\lambda(\pi-\xi_m')}}{\sqrt{x^2-(2m\pi)^2}}. \tag{5.8b}$$

Here the notation corresponds to that used for the respective poles and the factor $1/\pi$ standing in front of the integral is included. Provided λ is an even integer, the straight-line segments of the path of integration yield

$$-\frac{1}{2}\,\delta_{o\lambda}+\lim_{\rho\to 0}\frac{j}{\pi_\lambda}\left[2\int_\rho^\infty \frac{\cosh\lambda\eta\,d\eta}{e^x\sinh\eta-1}-P\int_\rho^{\pi/2}\cos\lambda\xi\,\cot\left(\frac{1}{2}\,x\sin\xi\right)d\xi\right], \tag{5.9}$$

where $\delta_{o\lambda}$ is the Kronecker delta. It should be noted that when the integrand in the finite integral is singular, the Cauchy principal value is to be taken, as indicated by the symbol P (the indicated limit as $\rho\to 0$ essentially is also a Cauchy value).

Upon comparing (5.4) and (5.6) it is seen that

$$\sigma_{pp}^{(1)}(ka)=2\,\mathrm{Im}\,[S_{2p}(2ka)-S_0(2ka)].$$

It happens that for $\lambda=2p$ (where $p=0, 1, 2, \ldots$) the net imaginary contribution of (5.8) to S_{2p} vanishes. Consequently (5.9) yields

$$\sigma_{pp}^{(1)}(ka)=\frac{8}{\pi}\int_0^\infty \frac{\sinh^2 p\eta d\eta}{e^{2ka\sinh\eta}-1}+\frac{4}{\pi}P\int_0^{\pi/2}\sin^2 p\xi\cot(ka\sin\xi)\,d\xi \tag{5.10}$$

wherein the integrands are now well-behaved at their lower limits and thus we were permitted to put $\rho=0$. The corresponding result for $\sigma_{pp}^{(2)}$ is

$$\sigma_{pp}^{(2)}(ka)=-\frac{8}{\pi}\int_0^\infty \frac{\sinh^2 p\eta d\eta}{e^{ka\sinh\eta}+1}-\frac{4}{\pi}P\int_0^{\pi/2}\sin^2 p\xi\tan\left(\frac{ka}{2}\sin\xi\right)d\xi, \tag{5.11}$$

which is easily obtained from (5.10) with the aid of (5.5).

(An interesting by-product of the above development, obtained by taking the real parts of (5.6), (5.8), and (5.9), is

$$\frac{1}{2}\delta_{0p}+\sum_{n=1}^{\infty}J_{2p}(nx)=\frac{1}{x}+2\sum_{m=1}^{Q}\frac{\cos\left[2p\sin^{-1}(2m\pi/x)\right]}{\sqrt{x^2-(2m\pi)^2}},$$

where $p=0, 1, 2, \ldots$; x and Q are as specified above. This result was obtained by Ignatowsky [12] by a different method.)

Thus far numerical results for the σ's have been obtained primarily from (5.10). The integrals were evaluated numerically by Paul F. Wacker and William W. Longley, Jr., using Legendre-Gaussian quadrature formulas programed for machine computation.[6] Values for $\sigma_{pp}^{(2)}$ were calculated from values of $\sigma_{pp}^{(1)}$ using (5.5). Results for the two lowest values of p are given in table 2. (More extensive tables are planned for the previously-mentioned separate publication.)

b. Conversion to More Rapidly Convergent Series

An alternative method of evaluating the σ's is afforded by a transformation of the Bessel-function series involved into more rapidly convergent series. This was accomplished a good many years ago by W. von Ignatowsky [12] in connection with his work on diffraction by gratings—a subject that will be recognized as being rather closely related to the subject of the present paper. Ignatowsky obtained

$$\pi\sum_{n=1}^{\infty}Y_0(nx)=\log\frac{2\kappa}{\gamma}+\sum_{n=1}^{Q}\frac{1}{n}+\sum_{n=Q+1}^{\infty}\left(\frac{1}{n}-\frac{\kappa}{\sqrt{n^2\kappa^2-1}}\right), \tag{5.12a}$$

$$\pi\sum_{n=1}^{\infty}Y_{2p}(nx)=\frac{1}{2p}-\frac{1}{2}\sum_{\mu=1}^{p}\frac{(2\kappa)^{2\mu}(p+\mu-1)!B_{2\mu}}{(2\mu)!(p-\mu)!}$$

$$-\kappa\sum_{n=1}^{Q}\frac{\sin\left[2p\sin^{-1}(n\kappa)\right]}{\sqrt{1-n^2\kappa^2}}-(-)^p\kappa\sum_{n=Q+1}^{\infty}\frac{\left(n\kappa-\sqrt{n^2\kappa^2-1}\right)^{2p}}{\sqrt{n^2\kappa^2-1}}. \tag{5.12b}$$

In these expressions $\kappa=2\pi/x$; $p=1,2,3, \ldots$; Q is as defined above; $\log\gamma=0.57721566\ldots$ is the Euler-Mascheroni constant; and the $B_{2\mu}$ are the Bernoulli numbers, here so labeled that $B_2=1/6$, $B_4=1/30$, $B_6=1/42$, $B_8=1/30$, $B_{10}=5/66$, etc. Equation (5.12a) is fairly well known at the present time. (A few numerical values are given for the series $\sum_{n=1}^{\infty}(-)^n Y_0(nx)$ in

[6] The estimates of accuracy of computed quantities given in this paper were made by Mr. Longley, who was responsible for most of the detail of the calculations.

[13].) The series for $\sigma_{11}^{(2)}$ obtained from (5.12a) and (5.12b) has been duplicated by an independent mode of derivation; this represents a partial check of the complicated expression (5.13b). Thus far in the present work, Ignatowsky's series have been used only to spot check some of the numerical results obtained from the definite integrals. The series are especially suitable for either estimation or calculation of the σ's when p is large.

The author has benefited from the discussion of many points with Paul F. Wacker and is indebted to David F. Wait for assistance with some of the theory and computations.

6. References

[1] D. S. Saxon, Notes on lectures by Julian Schwinger: Discontinuities in waveguide, Mass. Inst. Tech. Rad. Lab. report 43–7/4/44.
[2] F. E. Borgnis and C. H. Papas, Randwertprobleme der Mikrowellenphysik (Springer-Verlag, 1955).
[3] N. Marcuvitz, ed., Mass. Inst. Tech. Rad. Lab. Series **10,** 255–257 (McGraw-Hill Book Co., New York, N.Y. 1951).
[4] R. W. Beatty and D. M. Kerns, Recently developed microwave impedance standards and methods of measurement, IRE Trans. on Instrumentation **I–7,** 319 (1958).
[5] See [2], p. 24.
[6] R. Courant and D. Hilbert, Methods of mathematical physics, 1st Eng. Ed., p. 175 (Interscience, N.Y., 1953).
[7] F. Riesz – B. Sz.-Nagy, Functional analysis, p. 262 (Frederick Ungar Publishing Co., New York, N.Y., 1955).
[8] See [6], p. 474.
[9] P. M. Morse and H. Feshback, Methods of theoretical physics, p. 814 (McGraw-Hill Book Co., Inc., New York, N.Y., 1953).
[10] G. N. Watson, A treatise on the theory of Bessel functions, p. 361 (The University Press, Cambridge, England, 1944).
[11] See [6], p. 468.
[12] W. von Ignatowsky, Über Reihen mit Zylinderfunktionen nach dem Vielfachen des Argumentes, Arch. der Math. u. Phys. **23,** 193 (1914).
[13] V. G. Smith, L. Infeld, and W. J. Chien, On some series of Bessel functions, J. Math. and Phys. **20,** 22 (1947).
[14] P. Frank and R. v. Mises, Die Differential- und Integralgleichungen der Mechanik und Physik, Vol. 2, p. 808ff (Mary Rosenberg, New York, N.Y., 1943).

Boulder, Colo. (Paper 64B2–29)

Publications of the National Bureau of Standards*

(Including papers in outside journals)

Selected Abstracts

A comparison of experimental and theoretical relations between Young's modulus and the flexural and longitudinal resonance frequencies of uniform bars, S. Spinner, T. W. Reichard, and W. E. Tefft, *J. Research NBS* **64A,** *No. 2, 147 (1960).*

The relations between Young's modulus and the mechanical resonance flexural and longitudinal frequencies of two sets of steel bars has been established by an empirical method. The agreement between theoretical and experimental curves for flexural vibrations appeared somewhat better than that for longitudinal vibrations.

On the mode theory of very-low-frequency propagation in the presence of a transverse magnetic field, D. D. Crombie, *J. Research NBS* **64D,** *No. 3, 265 (1960).*

The effect of a purely transverse horizontal magnetic field on the propagation of VLF waves is considered. It is shown that the magnetic field introduces non-reciprocity, and that for propagation along the magnetic equator, the rate of attenuation is less for west-to-east propagation than for east-to-west propagation.

On the theory of reflection of low- and very-low-radiofrequency waves from the ionosphere, J. R. Johler and L. C. Walters, *J. Research NBS* **64D,** *No. 3, 269 (1960).*

The rigorous application of the magneto-ionic theory to the calculation of reflection coefficients for a sharply bounded ionosphere model is discussed. This is a comprehensive paper on the details which pertain to the rigorous evaluation of the reflection coefficients. The paper is illustrated with computations applicable to the *D*-region or the *E*-region of the ionosphere. The quasi-longitudinal approximation is derived from this theory and the range of validity of this approximation is illustrated. The restrictions imposed by the use of a sharply bounded model ionosphere are discussed.

Focusing, defocusing, and refraction in a circularly stratified atmosphere, K. Toman, *J. Research NBS* **64D,** *No. 3, 289 (1960).*

Focusing, defocusing, astronomical refraction and path length of rays as a function of the departure angle Δ of the ray at the source is described for cases with the source outside, inside, or on the boundary of a circular stratification. Relative to zero elevation angle symmetrical and centrosymmetrical distributions are found.

Impedance characteristics of a uniform current loop having a spherical core, S. Adachi, *J. Research NBS* **64D,** *No. 3, 297 (1960).*

The radiation impedance is derived by the EMF method in a convenient form as the sum of the self-radiation impedance of a loop in the free space and an additional term due to the reaction between the loop and the sphere which is proportional to the well-known expansion coefficient of a magnetic-type scattered wave from a sphere in an incident plane wave. The first antiresonance frequency has been given in the form of a universal curve for a very small uniform current loop with core of an arbitrary composition of μ_s and ϵ_s, subject to the condition that the refraction coefficient $N=\sqrt{\mu_s \epsilon_s}$ is extremely large. Some numerical calculations show that high-μ core is desirable for a comparatively lower frequency region, and high-ϵ core is rather desirable in an antiresonance region.

Basic theorems in matrix theory, M. Marcus, *NBS Applied Math. Series 57 (1960) 15 cents.*

This is a survey of the basic identities and inequalities of matrix theory. Included are results dealing with elementary properties, canonical forms, invariance, congruence, commutativity, orthogonalization, eigenvalues, determinants, submatrices, rank, determinant and rank inequalities, numerical methods for inversion and eigenvalues, condition numbers.

Relaxation processes in multistate systems, K. E. Shuler, *Phys. of Fluids* **2,** *No. 4, 442 (1959).*

An analysis is made of the relaxation of "multistate" systems, i.e., systems with more than two quantum states or two different chemical species, for linearized processes described by the "Master Equation" of the theory of transport processes. These results are compared with those obtained from the analysis of two-state relaxation processes and the concept of "relaxation time" is discussed in this framework. A discussion is presented of the transformation of microscopic to macroscopic relaxation equations. The existence of periodic and/or aperiodic oscillatory solution of the linear multistate relaxation equation is investigated. It is shown that the multistate relaxation equations admit of aperiodic oscillatory solutions.

Invariant and complete stress functions for general continua, C. Truesdell, *Arch. Rational Mech. Anal.* **1,** *No. 4, 1 (1959).*

This paper presents an organized survey of results, with proofs, concerning exact general solution of the underdetermined equations of motion of a continuous medium. Special attention is given to a flat space of n dimensions, to an n-dimensional space of constant curvature, and to a 2-dimensional surface. Two methods are used: (1) the representation of a solenoidal vector as the curl of a vector potential, and (2) a variational formulation, essentially the converse of the principle of virtual work.

The construction of Hadamard matrices, E. C. Dade and K. Goldberg, *Mich. Math. J.* **6,** *247 (1959).*

It is proved that a Hadamard matrix of order $4n$ can be constructed if there exists a transitive permutation group of odd order and degree $4n-1$ whose subgroups leaving one element fixed have three transitivity sets each.

Bending and stretching of corrugated diaphragms, R. F. Dressler, *ASME Trans.* **81D,** *No. 4, 651 (1959).*

Solutions of the exact linear elastic shell equations for all stresses and displacements in a typical corrugated diaphragm are presented for three cases over a 9 to 1 range of thicknesses. Results were obtained by numerical integration in an electronic digital computer. The effect of thickness variation is discussed with respect to both stresses and resultants, and peak values needed for design purposes are presented. Circumferential and meridional stresses are found to be equally important throughout the thickness range analyzed. Bending and membrane stresses are likewise equally important throughout the range. Peak values in some cases occur near the outer rim.

Modular forms whose coefficients possess multiplicative properties, M. Newman, *Ann. Math.* **70,** *478 (1959).*

Let q be a prime, r and s nonzero integers, and η (τ) the Dodekind modular form. All modular forms of type $\eta^r(\tau)\eta^s(q\tau)$ such that the related Dirichlet series possesses an Euler product are determined.

The analysis of latin squares with a certain type of row-column interaction, J. Mandel, *Technometrics* **1,** *No. 4, 379 (1959).*

A serious limitation in the use of latin squares is the confounding of the main effect of each factor with the interaction of the remaining two factors. In some cases, the interaction of rows and columns can be expressed as a multiplicative term of assigned factors associated with rows and columns. The analysis of such designs is presented in detail. An example is discussed. The method is extended to the case where the treatment effects are linearly related to a given concomitant variable. A brief discussion is given of the relation of the proposed method with other tests for nonadditivity.

A note on algebras, A. J. Goldman, *Am. Math. Mo.* **66,** *795 (1959).*

The usual sufficient conditions that both regular representations of an algebra A be faithful are shown not to be necessary. A necessary and sufficient condition (viz, that no nonzero member of A be either a left annihilator or a right annihilator) is derived.

Linear differential equations of the second order with a large parameter, F. W. J. Olver, *J. Soc. Indust. Appl. Math.* **7,** *306 (1959).*

An investigation is made of the differential equation

$$\frac{d^2w}{dz^2}=\{u^2p(Z)+q(Z)\}w$$

for large values of the parameter u.

List of Titles

Journal of Research, Section 64A, No. 2, March–April 1960. 70 cents.

A carbon-14 beta-ray standard, benzoic acid–7–C^{14} in toluene, for liquid scintillation counters. W. F. Marlow and R. W. Medlock.

A comparison of experimental and theoretical relations between Young's modulus and the flexural and longitudinal resonance frequencies of uniform bars. S. Spinner, T. W. Reichard, and W. E. Tefft. (See above abstracts.)

Determination of copolymer composition by combustion analysis for carbon and hydrogen. Lawrence A. Wood, Irving Madorsky, and Rolf A. Paulson.

Some effects of aging on the surface area of portland cement paste. C. M. Hunt, L. A. Tomes, and R. L. Blaine.

Conformations of the pyranoid sugars. I. Classification of conformers. H. S. Isbell and R. S. Tipson.

Tritium-labeled compounds III. Aldoses—*1*—*t*. H. S. Isbell, H. L. Frush, N. B. Holt, and J. D. Moyer.

Determination of aluminum in precipitation hardening stainless steel and high temperature alloys. Lawrence A. Machlan, John L. Hague, and Edward J. Meros.

Journal of Research, Section 64D, No. 3, May–June 1960. 70 cents.

A theory of radar scattering by the moon. T. B. A. Senior and K. M. Siegel.

A theory of wavelength dependence in ultrahigh frequency transhorizon propagation based on meteorological considerations. Ralph Bolgiano, Jr.

A preliminary study of radiometeorological effects on beyond-horizon propagation. F. Ikegami.

Trade-wind inversion as a transoceanic duct. M. Katzin, H. Pezzner, B. Y.–C. Koo, J. V. Larson, and J. C. Katzin.

An analysis of propagation measurements made at 418 megacycles per second well beyond the radio horizon (a digest). H. B. Janes, J. C. Stroud, and M. T. Decker.

On the calculation of the departures of radio wave bending from normal. B. R. Bean and E. J. Dutton.

On the mode theory of very-low-frequency propagation in the presence of a transverse magnetic field. D. D. Crombie. (See above abstracts.)

On the theory of reflection of low- and very-low-radiofrequency waves from the ionosphere. J. Ralph Johler and Lillie C. Walters. (See above abstracts.)

Focusing, defocusing, and refraction in a circularly stratified atmosphere. K. Toman. (See above abstracts.)

Response of a loaded electric dipole in an imperfectly conducting cylinder of finite length. Charles W. Harrison, Jr., and Ronald W. P. King.

Impedance characteristics of a uniform current loop having a spherical core. Saburo Adachi. (See above abstracts.)

Journal of Research, Section 64C, No. 2, April–June 1960. 75 cents.

Measurement of cobalt-60 and cesium-137 gamma rays with a free-air chamber. H. O. Wyckoff.

Apparatus for the measurement of the normal spectral emissivity in the infrared. Arthur G. Maki, Ralph Stair, and Russell G. Johnston.

Electrostatic deflection plates for cathode-ray tubes. I. Design of single-bend deflection plates with parallel entrance sections. II. Deflection defocusing distortion of single-bend deflection plates with parallel entrance sections. Lothar Frenkel.

The functional synthesis of linear plots. J. P. Vinti and R. F. Dressler.

Radiation field from a rectangular source. J. H. Hubbell, R. L. Bach, and J. C. Lamkin.

Microwave attenuation measurements with accuracies from 0.0001 to 0.06 decibel over a range of 0.01 to 50 decibels. G. F. Engen and R. W. Beatty.

Effect of oleophobic films on metal fatigue. H. E. Frankel, J. A. Bennett, and W. L. Holshouser.

Ratio-recording spectroradiometer. Harry K. Hammond III, Warren L. Holford, and Milton L. Kuder.

An intermittent-action camera with absolute time calibration. G. Hefley, R. H. Doherty, and E. L. Berger.

Screw-thread standards for federal services 1957, Amends in part H28 (1944) (and in part its 1950 Supplement), NBS Handb. H28 (1957)—Pt. II (1959) 75 cents.

Specification for dry cells and batteries (supersedes Circular 559), NBS Handb. 71, (1959) 25 cents.

Report of the 44th National Conference on Weights and Measures 1959, NBS Misc. Publ. 228, (1959) 65 cents.

Research Highlights of the National Bureau of Standards Annual Report, Fiscal Year 1959, NBS Misc. Publ. 229 (1959) 55 cents.

Distribution of mail by destination at the San Francisco, Los Angeles, and Baltimore Post Offices, N. C. Severo and A. E. Newman, NBS Tech. Note 27 (PB151386) (1959) $1.50.

A history of vertical-incidence ionosphere sounding at the National Bureau of Standards, S. C. Gladden, NBS Tech. Note 28 (PB151387) (1959) $2.00.

Photographic dosimetry at total exposure levels below 20 mr, M. Ehrlich and W. L. McLaughlin, NBS Tech. Note 29 (PB151388) (1959) 50 cents.

An atlas of oblique-incidence ionograms, V. Agy, K. Davies, and R. Salaman, NBS Tech. Note 31 (PB151390) (1959) $2.25.

Resistance diode bridge circuit for temperature control, L. H. Bennett and V. M. Johnson, NBS Tech. Note 34 (PB151393) (1959) 50 cents.

Service area of an airborne television station, M. T. Decker, NBS Tech. Note 35 (PB151394) (1959) 75 cents.

Rotational transfer in the fluorescence spectrum of OH($^2\Sigma^+$), T. Carrington, J. Chem. Phys. **31**, No. 5, 1418 (1959).

Determination of viscosity, J. F. Swindells, Ch. XII, Physical methods of organic chemistry, pt. I, p. 689 (Interscience Publishers Inc., New York, N.Y., 1959).

Triaxial tension at the head of a rapidly running crack in a plate, J. M. Frankland, Trans. ASME [E] **26,** No. 4, 570 (1959).

Poly(tetrafluorethylene)—a radiation-resistant polymer, L. A. Wall and R. E. Florin, Letter to Ed., J. App. Polymer Sci. **II,** No. 5, 251 (1959).

On the approximate daytime constancy of the absorption of radio waves in the lower ionosphere, S. Chapman and K. Davies, J. Atmospheric and Terrest. Phys. **13,** No. 1–2, 86 (1958).

An X-ray crystallographic study of Schroeckingerite and its dehydration product, D. K. Smith, Am. Mineralogist **44,** 1020 (1959).

A note on the scattering of electrons from atomic hydrogen, A. Temkin, Phys. Rev. **116,** No. 2, 358 (1959).

Design of free-air ionization chambers for the soft X-ray region (20–100 kv), V. H. Ritz, Radiology **73,** No. 6, 911 (1959).

Thermal conductivity of indium antimonide at low temperatures, E. V. Mielczarek and H. P. R. Frederikse, Phys. Rev. **115,** 888 (1959).

Model for vibrational relaxation of diatomic gases behind shock waves, R. Herman and R. J. Rubin, Phys. of Fluids **2,** No. 5, 547 (1959).

Absolute photometry of the aurora—I. The ionized molecular nitrogen emission and the oxygen green line in the dark atmosphere, M. H. Rees, J. Atmospheric and Terrest. Phys. **14,** 325 (1959).

Direction findings on whistlers, J. M. Watts, J. Geophys. Research **64,** No. 11, 2029 (1959).

Geomagnetic and ionospheric phenomena associated with nuclear explosions, S. Matsushita, Nature **184,** BA33 (1959).

Stress corrosion cracking in low carbon steel, H. L. Logan, Proc. Phys. Met. Stress Corrosion Fracture Conf. Pittsburgh, Pa., Apr. 2–3, 1959, reprinted from Met. Soc. Conf., p. 295 (Interscience Publishers, New York, N.Y., 1959).

Higher oxides of silver, H.E. Swanson, J. Electrochem. Soc. **106,** No. 12, 1073 (1959).

Oscillatory behavior of magnetic susceptibility and electronic conductivity, A. H. Kahn and H. P. R. Frederikse, vol. IX, Advances in Solid State Physics, p. 257 (Academic Press Inc., New York, N.Y., 1959).

Geomagnetic activity following large solar flares, C. S. Warwick and R. T. Hansen, J. Atmospheric and Terrest. Phys. **14,** 287 (1959).

Simplification of systems of units currently evolving in the electrical field, F. B. Silsbee, Systems of Units, Am. Assoc. Advance. Sci., p. 7 (1959).

Atoms and free radicals by γ-irradiation at 4.2° K, L. A. Wall, D. W. Brown, and R. E. Florin, J. Phys. Chem. **63**, 1762 (1959).

**Publications for which a price is indicated (except for NBS Technical Notes) are available only from the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C. (foreign postage, one-fourth additional). Technical Notes are available only from the Office of Technical Services, U.S. Department of Commerce, Washington 25, D.C. (Order by PB number). Reprints from outside journals and the NBS Journal of Research may often be obtained directly from the authors.*

UNIVERSITY OF MICHIGAN #56-160074
PHYSICS LIBRARY
2035 RANDALL LAB.
ANN ARBOR MICH
D-18220

UNITED STATES
GOVERNMENT PRINTING OFFICE
DIVISION OF PUBLIC DOCUMENTS
WASHINGTON 25, D.C.

OFFICIAL BUSINESS

PRIVATE USE TO AVOID
OF POSTAGE, $300
(GPO)

JOURNAL OF RESEARCH of the National Bureau of Standards Vol. 64B, No. 2 April–June 1960
B. Mathematics and Mathematical Physics

Contents

Page

Non-self-adjoint boundary value problems in ordinary differential equations. Werner Greub and Werner C. Rheinboldt. 83

Criteria for the existence and equioscillation of best Tchebycheff approximations. John R. Rice. 91

Note on the solution of Riccati's differential equation. H. Herbert Howe. 95

On a generalization of the index notation for absolute tensors of arbitrary order. Edmund H. Brown. 99

Upper and lower bounds for the center of flexure. Lawrence E. Payne. 105

Half-round inductive obstacles in rectangular waveguide. D. M. Kerns. 113

Publications of the National Bureau of Standards. 131

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington 25, D.C. Price 75 cents (single copy). Subscription price: $2.25 a year, $0.50 additional for foreign mailing

UNITED STATES GOVERNMENT PRINTING OFFICE, WASHINGTON : 1960